



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 105371

TO: Rebecca Prouty  
Location: CM1/10A13/10D01  
Art Unit: 1652  
Friday, October 10, 2003

Case Serial Number: 09/884889

From: Edward Hart  
Location: Biotech-Chem Library  
CM1-6B02  
Phone: 305-9203

edward.hart@uspto.gov

### Search Notes

Examiner Prouty,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

**THIS PAGE BLANK (USPTO)**

## STIC-Biotech/ChemLib

105371

**From:** Prouty, Rebecca  
**Sent:** Monday, October 06, 2003 8:48 AM  
**To:** STIC-Biotech/ChemLib  
**Subject:** Sequence Search

Art Unit 1652, 10A13  
Mailbox: 10D01  
308-4000  
Serial Number: 09/884,889

Please search and interference search SEQ NOS: 5-8

RECEIVED  
OCT - 6 2003  
STIC/BIOTECH DIV.  
(STIC)

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
GAL 6502 Tel: 305-9203

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 10/17/03  
Date Completed: 10/17/03  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

## TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

## VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 104/10  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**





# STIC SEARCH RESULTS

## Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor  
308-4258, CM1-1E01

## Voluntary Results Feedback Form

➤ I am an examiner in Workgroup:  Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature  
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 07:56:28 ; Search time 127.677 Seconds  
(without alignments)  
7819.780 Million cell updates/sec

## SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
	1	2262	100.0	2262	2	US-08-674-897A-5	Sequence 5, Appl1
	2	2282	100.0	2262	3	US-08-951-844-5	Sequence 5, Appl1
	3	2262	100.0	2262	4	US-09-412-347-5	Sequence 5, Appl1
	4	672.2	29.7	2238	2	US-08-674-897A-7	Sequence 7, Appl1
	5	672.2	29.7	2238	3	US-08-951-844-7	Sequence 7, Appl1
	6	672.2	29.7	2238	4	US-09-412-347-7	Sequence 7, Appl1
	7	618.6	27.3	2331	1	US-08-418-782-20	Sequence 20, Appl1
	8	618.6	27.3	2331	2	US-08-852-219-20	Sequence 20, Appl1
	9	618.6	27.3	430765	3	US-09-103-440A-2	Sequence 2, Appl1
	10	606.6	26.8	413329	3	US-08-228-662-1	Sequence 1, Appl1
	11	606.6	26.8	413329	4	US-08-228-662-1	Sequence 1, Appl1
	12	606.6	26.8	2235	1	US-08-418-782-21	Sequence 1, Appl1
	13	606.6	26.8	2235	2	US-08-852-219-1	Sequence 2, Appl1
	14	578.2	25.6	2221	1	US-08-418-782-22	Sequence 2, Appl1
	15	578.2	25.6	2221	2	US-08-228-662-2	Sequence 2, Appl1
	16	578.2	25.6	2221	1	US-08-852-219-2	Sequence 2, Appl1
	17	578.2	25.6	4795	2	US-08-313-185-45	Sequence 45, Appl1
	18	578.2	25.6	4795	3	US-08-032-644A-5	Sequence 45, Appl1
	19	577.6	25.5	4794	2	US-08-459-499-8	Sequence 8, Appl1
	20	514.8	22.8	2181	4	US-09-328-332-3989	Sequence 3989, Appl1
	21	137.2	6.1	620	2	US-08-757-653-148	Sequence 148, Appl1
	22	137.2	6.1	620	3	US-08-757-653-148	Sequence 148, Appl1
	23	137.2	6.1	620	2	US-08-757-653-148	Sequence 148, Appl1
	24	137.2	6.1	620	2	US-08-757-653-150	Sequence 150, Appl1
	25	137.2	6.1	620	4	US-08-520-946-148	Sequence 148, Appl1
	26	137.2	6.1	620	4	US-08-520-946-148	Sequence 148, Appl1
	27	137.2	6.1	620	4	US-08-520-946-148	Sequence 148, Appl1

```
Query Match      100.0%; Score 2262; DB 2; Length 2262;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2262; Conservative 0; Mismatches 0; Indels 0;
```

## ALIGNMENTS

RESULT 1  
US-08-674-887A-5  
Sequence 5, Application US/08674887A  
Patent No. 5939300  
GENERAL INFORMATION:  
APPLICANT: Robertson, Dan E.  
APPLICANT: Sanyal, Indrajit  
INVENTOR: Sanyal, Indrajit  
TITLE OF INVENTION: CATALASES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/674, 887A  
FILING DATE: 03-JUL-1996  
PRIORITY APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY AGENT INFORMATION: A:  
ATTORNEY NAME: PH  
REGISTRATION NUMBER: 38 347  
REFERENCE/DOCKET NUMBER: 09015/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5059  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE LENGTH: 2262 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
SEQUENCE: Coding Sequence  
LOCATION: 1...2269  
US-08-674-887A-5

Sequence 150, App  
Sequence 143, App  
Sequence 145, App  
Sequence 147, App  
Sequence 149, App  
Sequence 143, App  
Sequence 145, App  
Sequence 147, App  
Sequence 149, App  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 2, Appl

QY 1 ATGATATACCCATCCGCTGACGATCTACACAGTAGCTGCGACCAAGATGCGAGACATTT 60  
 DB 1 ATGATATACCCATCCGCTGACGATCTACACAGTAGCTGCGACCAAGATGCGAGACATTT 60  
 QY 61 GTTCCCTTGGTATCGCCAAAGGATAGAGCAATAAGGAGAGAGCTATGAGCGTAAATGT 120  
 DB 61 GTTCCCTTGGTATCGCCAAAGGATAGAGCAATAAGGAGAGAGCTATGAGCGTAAATGT 120  
 QY 121 CTGTGATACAGCGGTGGTATACCTGATCGATCGGTACTTCCCAAAAGATGTGGTGGCGAA 180  
 DB 121 CTGTGATACAGCGGTGGTATACCTGATCGATCGGTACTTCCCAAAAGATGTGGTGGCGAA 180  
 QY 181 GGGTGAACCTGGATATTTGCACTCAGCAAGATCGCAATCAGAACCGATGGATCGGAT 240  
 DB 181 GGGTGAACCTGGATATTTGCACTCAGCAAGATCGCAATCAGAACCGATGGATCGGAT 240  
 QY 241 TTCACTACCGGTGAGAGTAGGAGTAGCGAGCTCGATTTGCGAGGGGTGAGAGAATGTCAC 300  
 DB 241 TTCACTACCGGTGAGAGTAGGAGTAGCGAGCTCGATTTGCGAGGGGTGAGAGAATGTCAC 300  
 QY 301 CGGTTGATGACCGATAGCAAGAGTGTGGCCGCTGACTGTGGGCGCACTACGCGGGTTG 360  
 DB 301 CGGTTGATGACCGATAGCAAGAGTGTGGCCGCTGACTGTGGGCGCACTACGCGGGTTG 360  
 QY 361 ATGATCGATGCGATAGCCATCGCTGTGGCCGCTGACTGTGGGCGCACTACGCGGGTTG 420  
 DB 361 ATGATCGATGCGATAGCCATCGCTGTGGCCGCTGACTGTGGGCGCACTACGCGGGTTG 420  
 QY 421 GGTGTACCGGAGCCAGCGCTTTGCACCGCTCAACTCTGTGGCGGACAACTGACGCTG 480  
 DB 421 GGTGTACCGGAGCCAGCGCTTTGCACCGCTCAACTCTGTGGCGGACAACTGACGCTG 480  
 QY 481 GATAAAGCGCGCTCTCTGTGGCGGATCANGAAGAGTACGCGCAAAATATGATGCGG 540  
 DB 481 GATAAAGCGCGCTCTCTGTGGCGGATCANGAAGAGTACGCGCAAAATATGATGCGG 540  
 QY 541 CGAGACCTGATCTCTGGTGGCACCGTGGCTATGATGCTCATGSGGCTTACCTGCTTAC 600  
 DB 541 CGAGACCTGATCTCTGGTGGCACCGTGGCTATGATGCTCATGSGGCTTACCTGCTTAC 600  
 QY 601 GGCTTCTCTTTCGCGCGGTGCGATTTGGGAACCCGAAAAGATATCTACTGSGGTGAC 660  
 DB 601 GGCTTCTCTTTCGCGCGGTGCGATTTGGGAACCCGAAAAGATATCTACTGSGGTGAC 660  
 QY 661 GAAAAGAGTGTGCGACCTCTGTGACGACGCTACGCGAGCTGAGCAAGACGAGACG 720  
 DB 661 GAAAAGAGTGTGCGACCTCTGTGACGACGCTACGCGAGCTGAGCAAGACGAGACG 720  
 QY 721 ATGGAAAACCGCTGGCGGTGTCGAATTTGGGATACCGGCAAGAGATATCTAGCGGTGAT 780  
 DB 721 ATGGAAAACCGCTGGCGGTGTCGAATTTGGGATACCGGCAAGAGATATCTAGCGGTGAT 780  
 QY 781 ACGGCAACCCCTGATCCGCTGAGAACCGGACGACAGGTACTTGAACCTTTCGCCGATG 840  
 DB 781 ACGGCAACCCCTGATCCGCTGAGAACCGGACGACAGGTACTTGAACCTTTCGCCGATG 840  
 QY 841 CGGATGATGGCAATGCTCTGGGTGACGCGACGCGACGAGCTGTAACCTTTCGCCGATG 900  
 DB 841 CGGATGATGGCAATGCTCTGGGTGACGCGACGCGACGAGCTGTAACCTTTCGCCGATG 900  
 QY 901 CAGGTTAATGGCAATGCTCTGGGTGACGCGACGCGACGAGCTGTAACCTTTCGCCGATG 960  
 DB 901 CAGGTTAATGGCAATGCTCTGGGTGACGCGACGCGACGAGCTGTAACCTTTCGCCGATG 960  
 QY 961 CAGGCGCTTAGGTGGGCAACCCCAATGCAATGCAAGGCAAGCAAGCGCGTGCCTG 1020  
 DB 961 CAGGCGCTTAGGTGGGCAACCCCAATGCAATGCAAGGCAAGCAAGCGCGTGCCTG 1020  
 QY 1021 GGTATCGAAGGTGTGGAACCAACCAACCCCAAGAAATCGATATGGCTATTTTCGACCTG 1080  
 DB 1021 GGTATCGAAGGTGTGGAACCAACCAACCCCAAGAAATCGATATGGCTATTTTCGACCTG 1080

QY 1081 CTGTTGGCTACAATTGGGAACCTGAAAAGAGTCTGCGGTGCGCCACCACATTTGGGAACG 1140  
 DB 1081 CTGTTGGCTACAATTGGGAACCTGAAAAGAGTCTGCGGTGCGCCACCACATTTGGGAACG 1140  
 QY 1141 ATTGACATCAAAAACCAAAACATGCGGCTTACGCGCGAGGACGCTCTATTGCGCGAAC 1200  
 DB 1141 ATTGACATCAAAAACCAAAACATGCGGCTTACGCGCGAGGACGCTCTATTGCGCGAAC 1200  
 QY 1201 CCGATCATGACCGATCGCGATATGCGGATAAAGCTAAATCGACCTATCGCGCTATCTCG 1260  
 DB 1201 CCGATCATGACCGATCGCGATATGCGGATAAAGCTAAATCGACCTATCGCGCTATCTCG 1260  
 QY 1261 GAAATATCTATGGCGGCTCTGAGCTACTTCANGAAGAACTTTCGCGAAGCGGTGGTCAAG 1320  
 DB 1261 GAAATATCTATGGCGGCTCTGAGCTACTTCANGAAGAACTTTCGCGAAGCGGTGGTCAAG 1320  
 QY 1321 CTGAGCAACCGTGACTTGGCGCGGCAAAATCAAGTTACATCGGCGCGGAAGTGGCGCGAAG 1380  
 DB 1321 CTGAGCAACCGTGACTTGGCGCGGCAAAATCAAGTTACATCGGCGCGGAAGTGGCGCGAAG 1380  
 QY 1381 GACCTGATTTGGCAAGCAACCCGATTCGCGGAGCTACACGACTACTGCGAAGAGTGGTC 1440  
 DB 1381 GACCTGATTTGGCAAGCAACCCGATTCGCGGAGCTACACGACTACTGCGAAGAGTGGTC 1440  
 QY 1441 AAGCAGAAAATTCACAAAGTGGCTGTGACCTAGTAGTGTCTCCACCGCTTGGGAC 1500  
 DB 1441 AAGCAGAAAATTCACAAAGTGGCTGTGACCTAGTAGTGTCTCCACCGCTTGGGAC 1500  
 QY 1501 AGTCCCGCTACTTATTCGCGGTTCGATATGCGCGCGGTGTCTAACGTTGCCGCAATTCGC 1560  
 DB 1501 AGTCCCGCTACTTATTCGCGGTTCGATATGCGCGCGGTGTCTAACGTTGCCGCAATTCGC 1560  
 QY 1561 TTGGCCCCACAGCAAGTGGCAAGGCAACGCGCGAGCGCTTGGCAAAATGCTCGACG 1620  
 DB 1561 TTGGCCCCACAGCAAGTGGCAAGGCAACGCGCGAGCGCTTGGCAAAATGCTCGACG 1620  
 QY 1621 GTCTAGAGCAGATCTCTGCGACACACGGCGCTAGCATCGCGAGCTGCTGTCTGGCC 1680  
 DB 1621 GTCTAGAGCAGATCTCTGCGACACACGGCGCTAGCATCGCGAGCTGCTGTCTGGCC 1680  
 QY 1681 GTTACGGTACGCTATGAGAAAGCGCAAGACGACATTTACGATGTCCGCTTCCCTTC 1740  
 DB 1681 GTTACGGTACGCTATGAGAAAGCGCAAGACGACATTTACGATGTCCGCTTCCCTTC 1740  
 QY 1741 CTGAAGCGCTGGCGATCGGACCGCGAGATGACCGAGCAGACTCTTCCGACCGCTG 1800  
 DB 1741 CTGAAGCGCTGGCGATCGGACCGCGAGATGACCGAGCAGACTCTTCCGACCGCTG 1800  
 QY 1801 GAGCGCGCTGGCGATGGGTGTGCGCACTGGCAGAAAGATATTTGGTGAAGCGCGAA 1860  
 DB 1801 GAGCGCGCTGGCGATGGGTGTGCGCACTGGCAGAAAGATATTTGGTGAAGCGCGAA 1860  
 QY 1861 GAGATGCTCTGATCTGCGGACGCTGATGGGCTTAAAGCGCGGAAATGACCGTCTG 1920  
 DB 1861 GAGATGCTCTGATCTGCGGACGCTGATGGGCTTAAAGCGCGGAAATGACCGTCTG 1920  
 QY 1921 CTGGCGGATATCCGCTATGCGCAACCACTATGCTATGCTGCGCAACACGCGGCTATTCAC 1980  
 DB 1921 CTGGCGGATATCCGCTATGCGCAACCACTATGCTATGCTGCGCAACACGCGGCTATTCAC 1980  
 QY 1981 GATTGTGAGCGCAATTTGACCAAGCACTTTTTTGTGAACCTGACCGATATGGGGACAGC 2040  
 DB 1981 GATTGTGAGCGCAATTTGACCAAGCACTTTTTTGTGAACCTGACCGATATGGGGACAGC 2040  
 QY 2041 TGGAGCGGATAGGTAGCAACGCTTACGAATATCCGCAACCGGACCGGTCGCTGAG 2100  
 DB 2041 TGGAGCGGATAGGTAGCAACGCTTACGAATATCCGCAACCGGACCGGTCGCTGAG 2100  
 QY 2101 TGGAGCGGCTCTGGGCTGGATCTGATATTTGCTTCACTGCTACTGCTCTTACGCA 2160  
 DB 2101 TGGAGCGGCTCTGGGCTGGATCTGATATTTGCTTCACTGCTACTGCTCTTACGCA 2160  
 QY 2161 GAGTGTATCGCCGACGACGATATACGCGGAGAGCTTCTGTCAAGAGACTTGTGTCGCGCTGG 2220

```

Db 2161 GAAGTGTACGCCAGCATATACCGGAGAAAGTTGTCAGAGACTTTCGTGGCCGCTGG 2220
Oy 2221 ACCAAAGTATGACCGCCGCTTTGACGTCGCGGTAA 2262
Db 2221 ACCAAAGTATGACCGCCGCGTTCACGTCGCGTAA 2262

RESULT 2
US-09-884-5
? Sequence 1: Application US/08951844
? Patent No. 6074860
? GENERAL INFORMATION:
? APPLICANT: Robertson et al.
? TITLE OF INVENTION: Catalases
? NUMBER OF SEQUENCES: 8
? CONVENTION DATA:
? ADDRESS: CARMEL, BYRNE, RAIN, GILFILLAN,
? ADDRESSER: CECCHI, STEWART & OLSTEIN
? STREET: 6 BECKER FARM ROAD
? CITY: ROSELAND
? STATE: NEW JERSEY
? COUNTRY: USA
? ZIP: 07068
? COMPUTER FILE:
? MEDIUM TYPE: 3 1/2 INCH DISKETTE
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: WORD PERFECT 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/951,844
? FILING DATE: 08/435
? PRIORITY DATE: 08/435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/674,887
? FILING DATE: July 3, 1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Charles J. Herron
? REGISTRATION NUMBER: 28,019
? REFERENCE/DOCKET NUMBER: 331400-55
? TELEPHONE/TELEFAX INFORMATION:
? TELEPHONE: 201-994-1740
? TELEFAX: 201-994-1744
? INFORMATION FOR SEQ ID NO: 5:
? LENGTH: 2262 NUCLEOTIDES
? TYPE: NUCLEIC ACID
? TOPOLOGY: LINEAR
? MOLECULE TYPE: Oligonucleotide
US-08-951-844-5

Query Match 100.0%; Score 2262; DB 3; Length 2262;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ATGAATAACGATCCGCTGACGATCTACACAGTACTTGCAGCAAGCATGCAGGCACTTTT 60
Db 1 ATGAATAACGATCCGCTGACGATCTACACAGTACTTGCAGCAAGCATGCAGGCACTTTT 60
Oy 61 GTTCCCTTGGTATCCCAAGCGCATAGACCAATAAGGAGAGAGCTATGAGCGGTAATGT 120
Db 61 GTTCCCTTGGTATCCCAAGCGCATAGACCAATAAGGAGAGAGCTATGAGCGGTAATGT 120
Oy 121 CTTGTCTACGCAAGGATGACACCTGACCGTACTCTCCACAAAGATTTGGTCCGCGGA 180
Db 121 CTTGTCTACGCAAGGATGACACCTGACCGTACTCTCCACAAAGATTTGGTCCGCGGA 180
Oy 181 GGGTTGAACCTTGGATATTTTCATCAGCAAGATCGCAATCAGACCGCGATGCCGAT 240
Db 181 GGGTTGAACCTTGGATATTTTCATCAGCAAGATCGCAATCAGACCGCGATGCCGAT 240
Oy 241 TTCAACTCCCTGAGAAAGTACCGCAAGCTCGATTTTCGACCGCTGAGAAAGATGTCCAC 300

```

```

Db 241 TTCAACTCCCTGAGAAAGTACCGCAAGCTCGATTTTCGACCGCTGAGAAAGATGTCCAC 300
Oy 301 GGGTTGAACCTTGGATATTTTCATCAGCAAGTGGCCGCTGAGCGCACTTACGCGGTTTG 360
Db 301 GGGTTGAACCTTGGATATTTTCATCAGCAAGTGGCCGCTGAGCGCGCACTTACGCGGTTTG 360
Oy 361 ATGATCCGATATGGCTTGGCACTCCGCTGGCACTTACCGTATTCGTGATGGCGGTGGGGGC 420
Db 361 ATGATCCGATATGGCTTGGCACTCCGCTGGCACTTACCGTATTCGTGATGGCGGTGGGGGC 420
Oy 421 GGTGATCGGAGACGACGCGTGTGGACCGCTCAACTCACTCGCGCGCAAGCGTCACTCG 480
Db 421 GGTGATCGGAGACGACGCGTGTGGACCGCTTTCACCGCTCAACTCTCTGGCGGACAGCGTCACTCG 480
Oy 481 GATTAAGCGCCGCTGCTGTGCTGGCGATCAAGAGAGTACGGCAACAAAATCAGCTGG 540
Db 481 GATTAAGCGCCGCTGCTGTGCGCGGATCAAGAGAGTACGGCAACAAAATCAGCTGG 540
Oy 541 GCAGACCTGATGATCTGTGGCGACCGTGTATGAGTCCATGGCTTACTGGCTTAC 600
Db 541 GCAGACCTGATGATCTGTGGCGACCGTGTATGAGTCCATGGCTTACTGGCTTAC 600
Oy 601 GGTCTCTCTTTCGGCGCGCTCGATATTTGGGAACCGAAAAAGATATCTACTGGGGTGAC 660
Db 601 GGTCTCTCTTTCGGCGCGCTCGATATTTGGGAACCGAAAAAGATATCTACTGGGGTGAC 660
Oy 661 GAAAAGTGTGGCGCACCTCTGACGACGCTAGCGGACGTGACAGCAGAGACC 720
Db 661 GAAAAGTGTGGCGCACCTCTGACGACGCTAGCGGACGTGACAGCAGAGACC 720
Oy 721 ATGGAACACCCGCTGGCGCTGTCTCAAAATGGGTCTGATCTATGTGAACCGGAAGGTGT 780
Db 721 ATGGAACACCCGCTGGCGCTGTCTCAAAATGGGTCTGATCTATGTGAACCGGAAGGTGT 780
Oy 781 AACGGCCACCTGATCGCTGAGAGACCGCACAGCAGTACTTGAACCTTGCCCGTATG 840
Db 781 AACGGCCACCTGATCGCTGAGAGACCGCACAGCAGTACTTGAACCTTGCCCGTATG 840
Oy 841 GCGATGACGACGAAAAACCGACGCGCTCACAGCTGGCGGCCACACCGTGGTAAATGT 900
Db 841 GCGATGACGACGAAAAACCGACGCGCTCACAGCTGGCGGCCACACCGTGGTAAATGT 900
Oy 901 CACGTTAATGCAATGCTGCTGCTGTAGCCCTGACCCGACCAAGAGCTGTGAGTTGAAAC 960
Db 901 CACGTTAATGCAATGCTGCTGCTGTAGCCCTGACCCGACCAAGAGCTGTGAGTTGAAAC 960
Oy 961 CAGGGCTTAAAGTGTGGGCAACCCCAACATGAGAGAGGCAAGCAAGCGGTGACCTGG 1020
Db 961 CAGGGCTTAAAGTGTGGGCAACCCCAACATGAGAGAGGCAAGCAAGCGGTGACCTGG 1020
Oy 1021 GGTATTCGAAGTGTGTTGGACACCAACCCGACGAAATTCGATATGGGCTATTTGACACTG 1080
Db 1021 GGTATTCGAAGTGTGTTGGACACCAACCCGACGAAATTCGATATGGGCTATTTGACACTG 1080
Oy 1081 GTGTGGCTTACATTTGGAGCTGAAAGAGCTGCTCCGGTGGCCACATTTGGGAGCG 1140
Db 1081 GTGTGGCTTACATTTGGAGCTGAAAGAGCTGCTCCGGTGGCCACATTTGGGAGCG 1140
Oy 1141 ATTGACATCAAAAGAAAAAAGCGCGGTGACGCCGACGACCCCTCTATTGCCACAC 1200
Db 1141 ATTGACATCAAAAGAAAAAAGCGCGGTGACGCCGACGACCCCTCTATTGCCACAC 1200
Oy 1201 CGGATCATGCGATGCGATATGGGCAATAGAGTAATTCGACACTATGGGCTATTTGTC 1260
Db 1201 CGGATCATGCGATGCGATATGGGCAATAGAGTAATTCGACACTATGGGCTATTTGTC 1260
Oy 1261 GAAAATTTATGCGCGATTCCTGAGTACTTTCAAGAAAACTTTTGGCAGAGCGGTGCTCAAG 1320
Db 1261 GAAAATTTATGCGCGATTCCTGAGTACTTTCAAGAAAACTTTTGGCAGAGCGGTGCTCAAG 1320
Oy 1321 CTGAGCGACCGTGTGACCTGGGCCGGAATCAGTTACATCGCGCGCGGAAAGTGGCGGACAA 1380
Db 1321 CTGAGCGACCGTGTGACCTGGGCCGGAATCAGTTACATCGCGCGCGGAAAGTGGCGGACAA 1380

```

1321	CTGAGCGCAGCTGATCCCTGGGCCCCGAAATCAAGTACGTTACATCGCCCGGGAAGTCCGGCGAGAA	1381	1381	GACCTGATTTGGCCACAGACCCGATATGTCGGCAGGTAAACACCGACTACTGCCAGAAAGTGGTCC	1440	1440	1440
Db		Qy	Qy	1381	GACCTGATTTGGCCACAGACCCGATATGTCGGCAGGTAAACACCGACTACTGCCAGAAAGTGGTCC	Db	Db
		Qy	Qy	1381	GACCTGATTTGGCCACAGACCCGATATGTCGGCAGGTAAACACCGACTACTGCCAGAAAGTGGTCC	Db	Db
		Qy	Qy	1441	ACGCGAANAATTCGACAAAGTGGCGCTGAGCANTAGTGGAGATGGTGTCTCCACGCTTGGGAC	1500	1500
Db		Db	Db	1441	ACGCGAANAATTCGACAAAGTGGCGCTGAGCANTAGTGGAGATGGTGTCTCCACGCTTGGGAC	Db	Db
		Qy	Qy	1501	AGTGGCCGCTACTTATCGCGGCTTCCGATATGCGGGGTGGTCTATGCGGGTGTCTCCACGCTTGGGAC	1560	1560
Db		Db	Db	1501	AGTGGCCGCTACTTATCGCGGCTTCCGATATGCGGGGTGGTCTATGCGGGTGTCTCCACGCTTGGGAC	Db	Db
		Qy	Qy	1561	TTGGGCCCCACAGAACGAGTGGCAGGGCACAGAGTCGAGCCCTGGCGAAAGTGTCTGAGC	1620	1620
Db		Db	Db	1561	TTGGGCCCCACAGAACGAGTGGCAGGGCACAGAGTCGAGCCCTGGCGAAAGTGTCTGAGC	Db	Db
		Qy	Qy	1621	GTCTACGACGAGATCTCTCGCGACACCGGGGTGAGCATCGCGGAGCGTATGTTCTTGGCC	1680	1680
Db		Db	Db	1621	GTCTACGACGAGATCTCTCGCGACACCGGGGTGAGCATCGCGGAGCGTATGTTCTTGGCC	Db	Db
		Qy	Qy	1681	GGTACGCTAGGCTATCGAANAAGCCGGCAAGACGACGAGTGTACGATGTGGCGGTTCCTCTTC	1740	1740
Db		Db	Db	1681	GGTACGCTAGGCTATCGAANAAGCCGGCAAGACGACGAGTGTACGATGTGGCGGTTCCTCTTC	Db	Db
		Qy	Qy	1741	CTGAAAGCGTGGCTGGCATCTGCGACACCGGAGTAGCAGGTACGATACGATGCTTCCCTTC	1800	1800
Db		Db	Db	1741	CTGAAAGCGTGGCTGGCATCTGCGACACCGGAGTAGCAGGTACGATACGATGCTTCCCTTC	Db	Db
		Qy	Qy	1801	GAGCGCTGGCGCGTTCGCGCAACTGGCAGAGAAAGATGATGTGGTCACCAACACGGGCTATTCAAC	1860	1860
Db		Db	Db	1801	GAGCGCTGGCGCGTTCGCGCAACTGGCAGAGAAAGATGATGTGGTCACCAACACGGGCTATTCAAC	Db	Db
		Qy	Qy	1861	GAGATGCTGTGATGCTGGTGGCAGCTGATGAGTGGCTTAACCGCCCGGGAATGACCGTCTG	1920	1920
Db		Db	Db	1861	GAGATGCTGTGATGCTGGTGGCAGCTGATGAGTGGCTTAACCGCCCGGGAATGACCGTCTG	Db	Db
		Qy	Qy	1921	CTGGCGCGTATGCGGCTACTGTGGGCACCACTATGTGGCCACCAACACGGGCTATTCAAC	1980	1980
Db		Db	Db	1921	CTGGCGCGTATGCGGCTACTGTGGGCACCACTATGTGGCCACCAACACGGGCTATTCAAC	Db	Db
		Qy	Qy	1981	GATGCTGTGATGCGCAGTGTACCAACGACATCTTTTGTGATCTACCACTGATGGGACAGC	2040	2040
Db		Db	Db	1981	GATGCTGTGATGCGCAGTGTACCAACGACATCTTTTGTGATCTACCACTGATGGGACAGC	Db	Db
		Qy	Qy	2041	TGGAAGCCGGTAGTAGCAACGCCCTACGAATCCGGCAGCCGACAGCGGTGCCGTGAAG	2100	2100
Db		Db	Db	2041	TGGAAGCCGGTAGTAGCAACGCCCTACGAATCCGGCAGCCGACAGCGGTGCCGTGAAG	Db	Db
		Qy	Qy	2101	TGGACCGCCTCGGCGGTGGATCTGGTATTTGGTGTTCACATCGGTACTGGGGTGTAGCCA	2160	2160
Db		Db	Db	2101	TGGACCGCCTCGGCGGTGGATCTGGTATTTGGTGTTCACATCGGTACTGGGGTGTAGCCA	Db	Db
		Qy	Qy	2161	GAAGTGTAGCCCAAGAGATTAACCGCGAGAGTTGTGTAGAGACTTCTGTGGCGGCTCG	2220	2220
Db		Db	Db	2161	GAAGTGTAGCCCAAGAGATTAACCGCGAGAGTTGTGTAGAGACTTCTGTGGCGGCTCG	Db	Db
		Qy	Qy	2221	ACCAAGTGTAGTAACCGCGGACCGTGTTCGAGCGTGGTGTGTA	2282	2282
Db		Db	Db	2221	ACCAAGTGTAGTAACCGCGGACCGTGTTCGAGCGTGGTGTGTA	Db	Db

### RESULT 3

US-09-412-347-5

Sequence 5, Application US/09412347

; Patent No. 6410290

GENERAL INFORMATION:

APPLICANT: Robertson, Dan E.

; APPLICANT: Sanyal, Indrajit

APPLICANT: Adhikari, Robert

; TITLE OF INVENTION: CATALASE

; NUMBER OF SEQUENCES: 8

Query Match

Query match	100.00%	Score 2262	DB 4	Length 2262
Best Local Similarity	100.00%	Pred. No. 0		
Matches 2262	Conservative	Mismatches 0	Indels 0	Gaps 0

[illegible]

QY 1 ATGAATAACGCATCCGCTGACGATCTACACAGTAGCTTGCAGCAAGATGCAGAGCATTT 60

Db  
1 ATGAATAACGCATCCGCTGACGATCTACACAGTAGCTTGCAGCAAGATGCAGAGCATT 60

**Table 1**

61 GTTCCCTTGGTATCGCCCAAGGCATAGAGCAATAAGGGAGAGAGCTATGAGCGGTAATGT 120 QY

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

Db  
61 GTCCCTTGGTATCGCCCAAGGCATAGAGCAATAAGGGAGAGAGCTATGAGCGGTAATGT 120

[illegible]

QY 121 CCCTGTCATGCACGGTGGTAAACACCTCGACCGGTACTTCCAACAAGATTGGTGGCCGGA 180

D<sub>b</sub> 121 CCTGTTCATGTCACGGTGGTAACACCTCGACCGGTACTTCCACACACATTCCTCCCCCA 180

DD 121 CC TGTCATGGCACGGTGGTAACACCTCGACCGGTACTTCCACACAAGATTGGTGGCCGGAA 180

181 GGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAATCAGACCCGATGGATCCGGAT 240

QY 161 GGGTGGACCTGGATATTTTGCATCAGCAGAGATCGCATATCAGATCGGAT 240

Db 181 GGGTTGAACCTGGATATTTTGCATCAGCAAGATCGCAAAATCAGACCCGATGGATCCGGAT 240

02 000106ACCI080H1H11IGCAICAGCACAGGHCOCACHATCAGAACCCCATGGATCCCGAT 240

QY 241 TTCAACTACCGTGAAGAAGTACGCAAGCTCGATTTCGACCGCGTGAAGAAAGATGTCCAC 300

QY 481 GATTAAGCGCGCGCTGCTGCTGGCGGATGACAGAAAGTACGAGCAACAAATACAGCTG 540  
 Db 481 GATTAAGCGCGCGCTGCTGCTGGCGGATGACAGAAAGTACGAGCAACAAATACAGCTG 540  
 QY 541 CGAGACCTGATGATTCCTGGCTGGCAACCGCTGCTATGAGTCCATATGGGCTTACCTGGCTAC 600  
 Db 541 CGAGACCTGATGATTCCTGGCTGGCAACCGCTGCTATGAGTCCATATGGGCTTACCTGGCTAC 600  
 QY 601 GCGTCTCTCTTTTGGCGCGGTCGATATTTTGGGAACCGCAAAAGATATCTATCGGGTGAC 660  
 Db 601 GCGTCTCTCTTTTGGCGCGGTCGATATTTTGGGAACCGCAAAAGATATCTATCGGGTGAC 660  
 QY 661 GAAAAGAGTGTGTCACCTCTGACGAAGCTACAGGCGAGTGAACAAAGCCAGAGACC 720  
 Db 661 GAAAAGAGTGTGTCACCTCTGACGAAGCTACAGGCGAGTGAACAAAGCCAGAGACC 720  
 QY 721 ATGGAAGAACCCCTGCTGCGGCTGCGCAATTTGGATTTGATATGTGAAGCCGAGAGGTT 780  
 Db 721 ATGGAAGAACCCCTGCTGCGGCTGCGCAATTTGGATTTGATATGTGAAGCCGAGAGGTT 780  
 QY 781 AACGCCACCCCTGATCCCTGTAGAACCCGACAGAGGTACTTTGAACCTTCGCGCGCTATG 840  
 Db 781 AACGCCACCCCTGATCCCTGTAGAACCCGACAGAGGTACTTTGAACCTTCGCGCGCTATG 840  
 QY 841 CGATGATGACAGAAACCGCAACCGCCCTCAGACGCTGGGCGCAACCGCTGCTATGCT 900  
 Db 841 CGATGATGACAGAAACCGCAACCGCCCTCAGACGCTGGGCGCAACCGCTGCTATGCT 900  
 QY 901 CAGGTAATGGAACCTCTGCGTATAGCCCTGAGCCCAAAAGCCTCTGACGTTGAAAC 960  
 Db 901 CAGGTAATGGAACCTCTGCGTATAGCCCTGAGCCCAAAAGCCTCTGACGTTGAAAC 960  
 QY 961 CAGGCTGTAGTGTGGGCTGAGGCGCAACGAGCGCAGGCGCAACCGCTGACCTG 1020  
 Db 961 CAGGCTGTAGTGTGGGCTGAGGCGCAACGAGCGCAGGCGCAACCGCTGACCTG 1020  
 QY 1021 GGTATCGAGGTGCTTGGACACCAACCGCCAGCAAAATCGATATGGGCTATTTGACAGCTG 1080  
 Db 1021 GGTATCGAGGTGCTTGGACACCAACCGCCAGCAAAATCGATATGGGCTATTTGACAGCTG 1080  
 QY 1081 CTGTTGCGCTACAAATTTGGAACTGAAAGAGTCTGCGGCTGCCACCAATTGGGAACCG 1140  
 Db 1081 CTGTTGCGCTACAAATTTGGGAACCTGAAAGAGTCTGCGGCTGCCACCAATTGGGAACCG 1140  
 QY 1141 ATGACATCAAAAGGAACAAAGCGGCTGACGCCAGCGCCCTCTATCTGCGCAAC 1200  
 Db 1141 ATGACATCAAAAGGAACAAAGCGGCTGACGCCAGCGCCCTCTATCTGCGCAAC 1200  
 QY 1201 CGATCATGACGATGCGGATATGCGGATTAAGCTAAATCGACCTATGCGCTATCTGCG 1260  
 Db 1201 CGATCATGACGATGCGGATATGCGGATTAAGCTAAATCGACCTATGCGCTATCTGCG 1260  
 QY 1261 GAAATTTATGCGCGATCTGAGTACTTCAAGAAATTTTCGGAAGCGGTGGTCAAG 1320  
 Db 1261 GAAATTTATGCGCGATCTGAGTACTTCAAGAAATTTTCGGAAGCGGTGGTCAAG 1320  
 QY 1321 CTGAGCACCGCTGACTGGCGCGGAATCAGTTACATCGCGCGGAAGTGGCGCAGAA 1380  
 Db 1321 CTGAGCACCGCTGACTGGCGCGGAATCAGTTACATCGCGCGGAAGTGGCGCAGAA 1380  
 QY 1381 GACCTGATTTGGCAAGCCGATTCGCGCAGGTAAACCGACTACTCTGCGAAGAGTGGTC 1440  
 Db 1381 GACCTGATTTGGCAAGCCGATTCGCGCAGGTAAACCGACTACTCTGCGAAGAGTGGTC 1440  
 QY 1441 AACGAGAAATTTGCACAAAGTGCCCTGACAGTATGAGATGCTCTCCACCGCTGGGAC 1500  
 Db 1441 AACGAGAAATTTGCACAAAGTGCCCTGACAGTATGAGATGCTCTCCACCGCTGGGAC 1500  
 QY 1501 ATGCGCGCTACTTATCGCGGTTCGATATGCGCGCGGCTGCTCTAACGCTGCCCGCAATTCG 1560  
 Db 1501 ATGCGCGCTACTTATCGCGGTTCGATATGCGCGCGGCTGCTCTAACGCTGCCCGCAATTCG 1560  
 QY 1561 TTGGCGCCACAGAGAGTGTGTCAGGCGCAAGCGCGGAGCGCTGGCGAAGTGTGTCAGC 1620

## RESULT 4

US-08-674-887A-7

; Sequence 7, Application US/08674887A

; Patent No. 5939300

; GENERAL INFORMATION:

; APPLICANT: Robertson, Dan E.

; APPLICANT: Sanyal, Indrajit

; APPLICANT: Sanyal, Indrajit

; TITLE OF INVENTION: CATALASIS

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FastSoft for Windows Version 2.0

; CURRENT APPLICATION DATA: 08/674,887A

; FILING DATE: 03-JUL-1996





QY 1911 GACGCTGCTGGGGGGTATGGGGTACTGGGGACACAACTATGGTGGACACAAACACGG 1970  
 Db 1884 GACTGCTTGTGGGGGTATGGGGTACTGGGACACAACTACGCGGTGGGACGATGG 1943  
 QY 1971 GGTATTCACCGAATGTGAAGGCCAGTTGACACACGACTTTTGTGAACTCGACGATAT 2030  
 Db 1944 AGTGTTCACAAATGAAGCGGGTCACTATCCCATGACTCTTTGTAAACTGCTAGACT 2003  
 QY 2031 GGGGAACAGCTGG -----AAGCGGGTAGTAGAGCAAGCGCTACGAATTCGCGACCGCAA 2084  
 Db 2004 CAACTAATAATGGGAGCGGAGGATGATGACAGCAAAATTTTAGGAGAGATTCGA 2063  
 QY 2085 GACCGTGGCGTGAAGTGGACCGCTCGGGGTGGATCTGGTATTTGGTTCACATCGCT 2144  
 Db 2064 ACTGGCGAAGTGAAGTGGAGTGGACCGGAGACCTGATCTGGATCCAAATTCGA 2123  
 QY 2145 ACTGGCGCTTACGACAGAGTGAAGCCGACGACAGCAAGCAAGCGGAGAGTGTGACGA 2204  
 Db 2124 GCTAAGAGCCCTGGCAGAGTGTACGGCTGTGCGAGATTCGGAAGAAATTTGTTAAACA 2183  
 QY 2205 CTTCGTGGCGCTGGACCAAGTGAAGAGCGGACCGCTTCGA 2249  
 Db 2184 TTTTGTGAAGGCTTGGGCGCAAGTAATGAGCTGGACCGGGTTGA 2228

RESULT 5

US-08-951-844-7  
 : Sequence 7, Application US/08951844  
 : Patent No. 6074860  
 : GENERAL INFORMATION:  
 : APPLICANT: Robertson et al.  
 : TITLE OF INVENTION: Catalases  
 : PRIORITY: 08/07068  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: CABELLA, BYRNE, BAIN, GILFILLAN,  
 : STREET: 6 BECKER FARM ROAD  
 : CITY: ROSELAND  
 : STATE: NEW JERSEY  
 : COUNTRY: USA  
 : ZIP: 07068  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: 3.5 INCH DISKETTE  
 : COMPUTER: IBM PS/2  
 : OPERATING SYSTEM: MS-DOS  
 : SOFTWARE: WORD PERFECT 5.1  
 : CURRENT APPLICATION DATA:  
 : FILING DATE: 08/09/95, 844  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION NUMBER:  
 : APPLICATION NUMBER: 08/674,887  
 : FILING DATE: July 3, 1996  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Charles J. Herron  
 : FIRM: Herron, Smith & Associates, P.C.  
 : REFERENCE/DOCKET NUMBER: 331400-55  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 201-994-1700  
 : TELEFAX: 201-994-1744  
 : INFORMATION FOR SEQ ID NO: 7:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 2238 NUCLEOTIDES  
 : TYPE: NUCLEIC ACID  
 : TOPOLOGY: LINEAR  
 : MOLECULE TYPE: DNA

US-08-951-844-7  
 Query Match 29.7%; Score 672.2; DB 3;  
 Best Local Similarity 59.5%; Pred. No. 3.7e-179;  
 Matches 1277; Conservative 0; Mismatches 808; Indels 60; Gaps 6;

QY 150 CGGTACTTCCAAACAAAGATTGGTGGCCGAAGGTTGAACCTGGATATTTTGCATCAGCA 209  
 Db 99 GCGCACCACAAACAGGATTTGGTGGCCCAACATGCTCAACCTCGGCATCTTAGCCACACA 158  
 QY 210 AGATCGCAATCAGACCGGATGCGCGGATTTCACTACCTCGAGAGCTACGACAGCT 269  
 Db 159 TTCAATCGCTATGGGACCAACACGACCGGATTTTGACTATGCGGAGAGTTTAAAGCT 218  
 QY 270 GATTTTCAGCGCTGAGAAAGATGTCACGCGTTGATGACCGATAGCCAGAGATGGTG 329  
 Db 219 AGATCTGCGACGCTTAAAGGCGCTGCGAGCGCTAATGACAGTTTCAACAGCGCTGGT 278  
 QY 330 GCGGCTGATCTGGGCGCATCAACGCGGTTTGTATGATCGTATGGCTTGCATCTCGCTGG 389  
 Db 279 GCGACGAGATTAAGGCTATTAATGGCGCTTCTTATACGATGCGCTGGCAGACGCGCG 338  
 QY 390 CACCTACCGTATCTGTATGCGCGCTGGGCGGTGTACCGGAGCGAGCGCTTTGCACC 449  
 Db 339 CACCTACCGTATCTGTATGCGCGCTGGGCGGTGTACCGGAGCGAGCGCTTTGCACC 449  
 QY 450 GCTCAACTCTCGGCGGACACGCTCAGCTGATGAAGCGCGCTGCTGTGGCGCT 508  
 Db 399 TGTCAATAGTGGCGACAAATGCCAATCTGGATTAAGCAGCTTGTCTTTTGGCCAT 458  
 QY 510 CAGAAGAGTACGGCACAAATACATCGTGGCGACGCTGATGATCTGGCTGGCACCT 569  
 Db 459 CAACACAAATATCGGAAATATCTCTGGCGCGATCTATGATATCTACAGAAAGCT 518  
 QY 570 GCGCTATGATGCTCATGGCTTACCTGCTTACGGCTTCTCTTTGGCGCGCTGATATTTG 629  
 Db 519 AGCTCTGGAACATATGGGCTTTAAACCTTTTGGTTTTCAGGTTGGCAGACGATGATG 578  
 QY 630 GGAAACCGCAAAAGATATCTACTGCGGTGAGCAAAAGATGCTGTCGCACTTCTGACGA 689  
 Db 579 GGAAGCTTGAGAGATGATTAAGTCTACTGGAGAGAAACCGATGCTTGGCA 632  
 QY 690 AGCTGACGGGCGTGAACAAAGCAGACGATGAAACCGGCTGGCGCTGCTCAAAAT 749  
 Db 633 GCGCTATGAAGGTGACCGAGAGG-----TCGAAATCCCTGGGAGCGGTACAAAT 683  
 QY 750 GGGTCTGATCTATGTGAACCGGAGAGTGTATAGGGCCACCTGATGCTGTGAGACCGG 809  
 Db 684 GGAAGCTATCTATGAACCCCGAAGGCCCAACGCGCAACGCTATGCTGTGCTG 743  
 QY 810 ACACAGAGTACTTGAACACTTTCGCGCGTATGCGGATGAACGACGAGAAACCGACGCT 869  
 Db 740 GCGGTATATTCGTGAGACTTTTGGCGCAATGGCAATGAATGACAGAAACCGGCTCT 803  
 QY 874 CAGAGCTGGGCGACAGCGCTGCGGTATGTGAGGTATGCGAGGTATGCGAGAGCTCTGCT 863  
 Db 804 CATAGCGGTGAGACACACCTTCGGAACCAACATGCTGCTGCGGTGGGAGAAATGT 863  
 QY 927 AGCCCTCGACCCCAAAAGGCTCTGAGGTGAAACACGAGGGTTAGTTGGGCAACCCCA 986  
 Db 864 GGGCGGACAGCTCTCGCGCGCAGTATGAGAAATGAGCTGGGTGGAAACCACTA 923  
 QY 987 CATCGAGGCGACGAGCAGCGCTGCGCTGCGGTATGAAAGTCTGTGGACACCA 1046  
 Db 924 CGGCGCGGACGAGGTGGGATACCATCAGCAATGATGAGTGAAGGCGCTTGACCAAC 983  
 QY 1047 CCCCACAGAAATTCGATATGGGCTATTTTCGACCTGCTGTGGCTACAAATTTGGAACTGA 1106  
 Db 984 CCCTACTCAATGGAGCAATTAACCTTTTGAACACCTCTTTGGTACGAGTGGAGCTTAC 1043  
 QY 1107 AAGAGTCTGCGGCTGCCACCATGTTGGGAGCGGATGACATCATCAAAAAGGAAACAGCG 1166  
 Db 1044 CAAGAGTCCAGCTTGAAGCTTATCTGGAACCAAGAGGCGGTGGCGGCTGGCGAT 1103  
 QY 1167 GGTGAGCCGACGAGCCCTCTATTCGCCAACAACCGGATCATGACGATGGGATATGCG 1226  
 Db 1104 ACGGGATGCAATATCCCAACAGTGGCAGCCTCCATTTATGCTCTCATCTAGGAGCTGGC 1163  
 QY 1227 GATTAAGGTAAATCGGACCTATCGCGCTATCTGCGGAAAAATCATGCGCGGATCCTGAGTA 1286

```

Db 1164 GCTGGCAGTGGACCTGATTACCAAAATATTTCTGACGGTACTATGAACACCTGATGA 1223
QY 1287 CTTCAAGAAATCTTCCGAAGCGGTGTCAAGCTGACGACCGGTGACCTGGSCCGAA 1346
Db 1165 TTTTCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1223
Db 1224 GTTTCGAGATGCTTTCGGAAGAGCATGTGTAACAAGTACACACAGCATATGGACACAA 1283
Db 1347 ATCACTGTTAGATCGGCCGGAAGTGGCGGCAAGAGCATGTTTGGCAAGACCGGATGCC 1406
QY 1284 GTTCGGCTACTGGGACACCAAGATGGCTCGAGGAGCACTCATCTGCGCAAGCACTTAC 1343
Db 1407 GGCAGCTAC-----ACCAGCATCTCGGAAGAGTGGTCCAGCAAGCAATGCC 1454
QY 1344 AGATGTAGCCATCTCTCTGTGACGAAACAGATATGAAGCGCTTAAAGCCAAATCCT 1403
Db 1455 ACAAGTGGCCGTAGCATAGTGAAGTGGTCTCCACCGCTTGGACAGTCCCGCTACTTA 1514
QY 1404 GBATCGGGAGTACAGGTAGGAGCTGGTGAAGCAGCGCATGGGCTTCGCTACTACTT 1463
Db 1515 TCCGGSTTCCGATATCGCCGGGGTGTACCGTGGCCGCATTCGCTTGGCCGCCACAGAA 1574
Db 1464 TAGAAATCTTGACACACCGCGGGGTGCCACAGCTGCACATATACGACTGGCCGCCACAAA 1523
QY 1575 CGATGGGAGCGAAGCGCGGAGCGCGCTGGCGAAAGTGGTG----- 1619
Db 1524 ASATCGGGATGACCAACCTCTGACAGCTTCCAGGGTACTTAAACACTAGAGATAT 1583
QY 1620 -----CGTCTACGAGCAGATCTTCCGACACCGCGCTAGCATCCGCGACTGAT 1670
Db 1584 CCAGGAGGACTTTTAAACGAGCGCAATCAGATACAAAGCAGTATCGTTGGCGGACCTGAT 1643
QY 1671 CGTCTGGCGGTAGCTGAGGATCAGAAAGCGCGCAAGCAGCAGATTTAGCATGTGGC 1730
Db 1644 TGTGCTGGCGGTGTGGCGGTGTAAAGAAAGCTTCCAAAGAGTCTGGCCATGAGTGCA 1703
QY 1731 CGTTCCTCTTCAAAAGCGGTGGCGATGCGACCGCGAGATGACCGACGACGACTCCTT 1790
Db 1704 GTGCGCTTTCAACCCGCGGAGCGGATGCCACCGCTGAGCAACCGCATGTGGAGCTTT 1763
QY 1791 GCGACCGCTGGCGGTGGCGATGGTTCGCACTGGCGAGAGAGTATGTGT 1850
Db 1764 CGAGCGCTAGACGCGCTGACGGCTTTAGAACTATCAATCAACCGGACGATCAAGT 1823
QY 1851 GAAGCGGAGAGATGCTGTGGATCGTGCACGATGTGGGCTTAACCGCGCGGAAT 1910
Db 1824 ATCCGCTGAGAAATGCTGTGAGACGGCGGCGCATCTGTGCGCTTTGGCGACAGAAAT 1883
QY 1911 GACCTGCTGTGGCGGTATGCGGTATGCGGAGCGCAACATGTGTGGCAACACAGCG 1970
Db 1884 GACTGCTTGTGAGCGGTATGCGGTACTGGGCAACCACTACGACGGTTCCGACGATG 1943
QY 1971 CGTATTCACCGATGTGAAGGCGAGTTGAACAACGACTTTTTTGTGAACCTGACCGATAT 2030
Db 1944 AGTGTGTTACMAATGCTGTGAGACGCTGACGATTCATCAATGACTTCTTGTAAACCTGTAGACT 2003
QY 2031 GGGGACCACTGG-----AAGCGGTAGTGAAGCAAGCGCTACGAAATGGCGGACGAA 2084
Db 2004 CAACACTAATGGCGAGCGAGTGATGAATCAACGAAAGTTTTTGAAGCGAGAGCTTCCA 2063
QY 2085 GACCGGTGGCTGAGTGGACCGCTCGCGGGTGGATGTGTTATTTGGTTCCACTCGCT 2144
Db 2064 ANCTGGGAGGTAAAGTGGAGTGGACCGCGGGTAGCCTGATCTTCGATTCMAATTCGA 2123
QY 2145 ACTGGCTCTTACGAGAGTGTACGCCAGAGAGATACCGGAGAGTGTGTGTCGATGA 2204
Db 2124 GCTAAGAGCGCTTGGCAGAGTGTACGGCTGTGCGAGTCTCTGAAGAAAGTTTGTTAAGA 2183
QY 2205 CTTGCGCGCGCTGGACCAAAAGTGAATGACGCGACCGCTTTCGA 2249
Db 2184 TTTTGTGAAGCGCTTGGCGAAAGTAAAGGACTTGGACCGGGTTGA 2228

```

RESULT 6

```

US-09-412-347-7
; Sequence 7, Application US/09412347
; Feature No. 412347
; GENERAL INFORMATION:
; APPLICANT: Robertson, Dan E.
; APPLICANT: Sanyal, Indrajit
; APPLICANT: Adhikari, Robert S.
; TITLE OF INVENTION: CATALASES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; STRESSES: Ish & Richardson P.C. 1400
; STREET: 4235 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: PASTISO for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/412,347
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/674,887
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09015/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; FAX: 619/678-5069
; INFORMATION FOR SEQ ID NO. 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE KEY:
; NAME/KEY: Coding Sequence
; LOCATION: 1...2235
US-09-412-347-7

```

Query Match 39.7% Score 672.2; Db 4; Length 2238;  
 Best Local Similarity 59.5%; Pred. No. 3.7e-179;  
 Matches 127; Conservation 0; Mismatches 808; Indels 60; Gaps 6;

```

QY 150 CGGTACTTCCACAAAGATTTGGTGGCGGAAGGGTTGAACCTGCGATATTTTCATCACA 209
Db 99 CGCGACCAAAACAGGGGATTTGGTGGCGCACATGCTCAACCTCGCATCTTACGCCACA 158
QY 210 AGATCGCAATCAGACCGATGGATTCGATTCACCTACCGTGAAGAGTACCGAAGCT 269
Db 159 TCTATCGCTTCTGGACCAACGACCGGATTTTGTCTATCCGAGAGTTTGAAGAAGCT 218
QY 270 GCAATTCGACGCGCTGAAGAAAGATGTCCAGCGTGTGATCGAGTACGACGATAGCCAGAGTGGTG 329
Db 219 AGATCGGCAAGCGGTATAAAGAGACCTTGGCAGCGCTAATGACAGATTCACAGGACTGGTG 278
QY 330 GCGCGCTGACTGGGCGCATACCGCGGTTTGATGATCCGTATGGCATTCGCGACTCCGCTGG 389
Db 279 GCGACGATATAGCTGTATTTGGCCCTTCTTTATACGATATGCGTGGCGACGCGCG 338
QY 390 CACCTACCGTATGCTGATGGCGCGTGGCGGTGATACCGGAAGCCAGCGCTTTCACCC 449
Db 339 CACCTACCGTATGCTGATGGCGCGTGGCGGTGATGGCGGTGCGTGGCGGTGATGACGCGCTTCGCGCC 398
QY 450 GCTCACTCTGCGCGGACGCTGACGCTGATGAAGAGCGGCGCTGTGCTGTGGCGCAT 509
Db 399 TCTCAATAGCTGGCAGACATGCCAATCTGGAATAGGCAAGCTGTGCTTCTTGGCCAT 458

```



APPLICATION NUMBER: US/08/418,782

FILING DATE: 04/02/2003

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Koesner, Warren D.

REGISTRATION NUMBER: 30,440

REFERENCE/DOCKET NUMBER: 150,1410S1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-339-0331

FAX: 612-339-3061

INVENTOR INFORMATION:

SEQUENCE CHARACTERISTICS:

LENGTH: 2331 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

FEATURES:

NAME/KEY: CDS

LOCATION: 70..2289

US-08-418-782-20

Query Match 27.3%; Score 618.6; Db 1; Length 2331;  
Best Local Similarity 58.7%; Num. Mismatches 819; Gaps 8;  
Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;  
160 AACAGAGATTGGCGGAGAGGTTGAACCTGGATATTTTGCATCAGCAAGTCGCA 219  
172 AACACAGGACTTGGTGGCCACACGGCTCACTCTGAAGTACTGCAAAAACCCGGCGTC 231  
220 TACAGCCGATGGATCCGGATTTTCAACTACCGTGAAGAGTACGCAAGCTGATTCGAC 279  
232 GCTGCCCGATGGGTGGCGCGCTTGCGATTCGCGGAGAGTCCGACATCGACGTTGAC 291  
280 GCCTGACGAAGAAGATGTCACGCGTGTATGCGGATGACGCAAGTGGTGGCGCTGAC 339  
292 GCCCTGACGGGGACATCGAGGAAGTGTATGACCACTCGCAGCCGCTGGTGGCCCGCAG 351  
340 TGGGGGCACTACGGCGGTTTGATGATCCGTATGGCTTGGCACTCCGCTGGCACTACCT 399  
352 TAGGGCACTTAGGGCGCTGTTTATCGGATGGGCTGGCACTCCGCACTTCGCC 411  
400 ATTGCTGATGGCGGTGGGGCGGTGTACCGGAAGCAGCGCTTTCACCGCTCACTCC 459  
412 ATCCAGCGACCGCGCGCGCGCGCGGGCGGCGGATGACGCGCTTCGCGCGCTTACAGC 471  
470 TGGCGCAAGCACTGACCTGATTAAGAGCGCGCGCTGCTGTGGCCGATCAAGAGAG 519  
462 TGGCGCAAGCACTGACCTGATTAAGAGCGCGCGCTGCTGTGGCCGATCAAGAGAG 531  
520 TAGCGCAAGCACTGACCTGATTAAGAGCGCGCGCTGCTGTGGCCGATCAAGAGAG 579  
532 TAGCGCAAGCACTGATTAAGAGCGCGCGCTGCTGTGGCCGATCAAGAGAG 591  
580 TCGAGCGCGCTTACCTGCTAGCGCTTCTTTCGCGCGCTGCTGATTTGGAGACCGAA 639  
592 TCGATGGGCTTCAAGAGCTCGGGTTCGGCTTCGGCGGGCTGCGATGGGCGCG-- 649  
640 AAGAGATATCTACTGGGCTGAGCAAAAAGATGGCTGGCACTTCTGACGAAGCTTACGGC 699  
650 -ATAGGCTCTATTGGGGCAAGGACCACTGGCTCG-----GCCATGAGCTTTACAGC 702  
700 GAGGTGACACCGCAGAGCCATGGAACACCGCTGGGGCTGTCCAAATGGGTGATC 759  
703 GGTAAAGCGGCTC-----TGGACMACCCGCTGGCGGCTGCGATGGGCTGATC 753  
760 TAGTGTGAACCCGAGAGTGTAAAGAGTGTACCGCTGGGGCTGTCCAAATGGGTGATC 799  
754 TAGTGTGAACCCGAGAGCGAGACGACCAACCGGACCCCATGGCGGGCGGTGCACTT 813  
820 CTGAAAGCTTCGCGCTGATGGGCTGATGAGCAAGCAAGCAAGCAAGCAAGCAAGCA 879  
814 CGGAGAGCTTTCGGCGCATGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 873

880 GGCACACACCGCTCGGTAAATGTTTCACGTAATGCCATGCTCTGCGTTAGCCCCCTGACCA 933  
874 GGTCACTTCTTGGTAAAGCCATGCGCGCGCGCGCGCTGCTGTGGCGCCGCAACCC 933  
940 AAGAGCTCTGACGTTTGAANAACAGGGCTTGTAGTTGGGCAACCCCAACATGACGAGCAG 999  
934 GAGGCTGCTCCGCTGGAGAGATGGGCTTGGGCTGGAAGAGCTGATGAGACCGGAAC 993  
1000 GCAAGCAACCGCTGACCTCGGGTATGGAAGGTGCTGGACACCAACCCAGCAATTC 1059  
994 GGTAAAGCGCATGACAGCGGATGAGTGTATGGAACACACCGCGAGAAATTC 1053  
1060 GATATGGCTATTTTCAGACTGCTGTCTGGCTACAAATTTGGGCACTGAAAGAGCTCTCGC 1119  
1054 GACACACATTTTCTCGAGATCTCTGACGGCTACGAGTGGGAGCTGACGAGAGCCCTGCT 1113  
1120 GTGTGCCACCATTTGGGACCGATTTGACATCAAAAAGAAACAGACCGGTTGACGCCAGC 1179  
1114 GGGGCTTGGCATACACGCCAAGAGCGGCGCGCTGCCGCCACATCCCGACCGCTTC 1173  
1180 GACCCCTCTATTGGCCGACACACCCGATCATGACGATGCGGATATGAGCAATAAGGTAAAT 1239  
1174 GCGGGGCC--AGGGCGCTTCCCGAGCATGTGGCCACTGACCTCTCGCTGGGTGGAT 1320  
1231 CGGATATGAGCGGATCAAGCGCTGCTGGTGGACACCCGAGGATTCGCGGAGAG 1390  
1240 CGCACTATGCGCGCTATTCGCGGAAAAATTCATGCGCGATCTCTGATCTCAAGAAACT 1399  
1300 TTGCGAAGAGCGTGGTTTCAAGCTGACACCGTGAACCTGGCCCGGAAATCAGTGTATCATC 1359  
1291 TTGCGCAAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1350  
1360 GCGCGGGAATGCGCGAGAGAGCTGATTTGCGAAGCCGATTTGCGGCAAGTGAACAC 1419  
1351 GGGCGCTGCTCCCAAGACAGACCTGCTGTGGAGAGTCCGCTCTCTGGGTGAGGCA 1410  
1420 GACTACTGCGAAGAGAGTGG-----TCAAGCAGAAAAATTCACAAAGTGGCGCTG 1467  
1411 GACCTGTCGGGGAAGCGAGATTCGACGCTTAAAGCGCGATCGGGCATCGGGATTC 1470  
1468 GACATTAAGTGAAGTGTCTGACCGCTTGGCAAGTGGCGCGCTTACTTATCGCGTTCGAT 1527  
1471 ACTGCTCACAGTAGTTTGCACCGCATGGCGGCGCGCTGCTGCTGCTGCTGATGAC 1530  
1528 ATGCGCGCGGTGGCTTAAAGGCTGCGCGATGCTGGCTTGGCGCCCAAGAGAGTGGCAGCGC 1587  
1531 AAGCG 1590  
1588 AAGAGAGCGGA---GGCGCTGGCGAAGAGTGTGCGCGCTACGAGCAAGATCT----- 1636  
1591 AAGGACCGCGAGGGGATTCGCGCAAGGTCATTCGACCTCGAAGAGATCCAGGATGCA 1650  
1637 -----CTGCGCAACACCGCGCTAGCATCGCGCATGATGCTGCTGCTGCTGCTGCTG 1680  
1651 TTCATCTCCG 1710  
1681 GGTAGCGTGGGCTGAGAGAGCG 1740  
1711 GCGTGTGCG 1770  
1741 CTGAAGAGCGCGTGGCGATGCG 1800  
1771 ACCCG 1830  
1801 GAGCGCGTGGCG 1860  
1831 GAGCGCGAGAGTGGCTTCGGAACACTACCTCGGAAGGCGAACCGCTTCCGCGCGCGAG 1890  
1861 GAGATGCTGCTGGATGCTGGCGACGCTGATGGGCTTAACCGCGCGCGGAGTAAGCTGCTGCTG 1920  
1891 TACTGCTGCTGCAAGGCGAACCTGCTTACGCTCAGTGGCGCGCGCGCGCGCGCGCGCGCG 1950

Qy 1921 CTGGCGGTGATGGCGTACTGGGCAACACACTATGAGTGGCAACAAACCGGGTATTACAC 1980  
 Db 1951 GTAGTGGCTGCGGCTCTGGCGCAATCAACGCTTACGGCTGGGGTGTACCC 2010  
 Qy 1981 GATGTGAAGGCCAGTTGACCAACGACTTTTTTGTGAAGCTGACCATATGGGAACAC 2040  
 Db 2011 GAGCGCTCCGAGTCACTGACCAACGACTTCTTCGTGAAGCTGCTGCAATGGGTATCAC 2070  
 Qy 2041 TGGAGC---GGTAGTAGACACGCTAGCAATCTCCGACCGAAGACGGTCCGCG 2097  
 Db 2071 TGGAGCGCTCGCAGTAGACGAGGACTTACCAAGGCAAGTAGCGATGGCAAGGTG 2130  
 Qy 2098 AAGTGAAGCGCTCTCGGGGTGGATCTGTATTTGGTTCGAAGTGTCTATCGCGCTTTAC 2157  
 Db 2131 AAGTGAAGCGAGCGACCGCTGGAGCTGTCTTCGGGTCCAACTCGAGGTTCGGGCGCTT 2190  
 Qy 2158 GCACAGTAGACCGACGAGATACCGGGAGAGTCTGTACAGACTGCTGCGCGCG 2217  
 Db 2191 GTGAGGTCTATGGCGCGGACGACGACGAGGAGTCTGTGCAAGACTTCGTGCGAGCT 2250  
 Qy 2218 TGGACCAAGTAGTAACCGCGACGCTTTCGAGCT 2252  
 Db 2251 TGGACAAAGTAGTAACCTCGACAGGTTGCGAGCT 2285

## RESULT 8

US-08-452-219-20  
 ; Sequence 20, Application US/08852219  
 ; Patent No. 5922575  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cockerill, Franklin R.  
 ; APPLICANT: Kline, Bruce C.  
 ; INVENTOR: Kline, Bruce C.  
 ; TITLE OF INVENTION: Detection of Isopiazid Resistant Strains  
 ; TITLE OF INVENTION: Detection of M. tuberculosis  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Muehling, Raasch & Gebhardt, P.A.  
 ; STREET: 119 No. 5922575th Fourth Street, Ste. 203  
 ; CITY: Minneapolis  
 ; STATE: MN  
 ; COUNTRY: USA  
 ; ZIP: 55401  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PC-DOS/MS-DOS  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/452,219  
 ; FILING DATE: 07-May-1997  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Sandberg, Victoria A.  
 ; REGISTRATION NUMBER: 41,287  
 ; ADDRESS: PO BOX 10000, MNK, 230.00010130  
 ; TELECOMMUNICATIONS INFORMATION:  
 ; TELEPHONE: 612-305-1226  
 ; TELEFAX: 612-305-1228  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2331 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; ORIENTATION: 5' to 3'  
 ; MOLECULE TYPE: DNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 70..2289  
 ; US-08-452-219-20

Query Match 27.3%; Score 618.6; DB 2; Length 2331;  
 Best Local Similarity 58.7%; Fred. No. 4.6e-164;

Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;  
 Qy 160 AACAAAGATGTGGTGGCGGAGGTGTGAACCTGGATATTTTGCATCAGCAAGATCCGAA 219  
 Db 172 AACCAAGACTGGTGGCGCCCAACGGCTCAATCTGAAGTACTGCAACAAACCGCGCTC 231  
 Qy 220 TCAGACCCGATGATGCGGATCTCAACTACCGTGAAGAAGTACGCAAGCTGGATTTGAC 279  
 Db 232 GCTGAACCGATGATGGTGGCGGCTTCGACTATTCGCGGAGGTTCGACATCATCGAGCTG 291  
 Qy 280 GGGCTGAGAAAGATGTCCCAACGGTGTGATCCAGTACGCAAGATGSGTGGCGGAGAC 339  
 Db 292 GCGCTGAGCGGCATTCGAGAAAGTATGACCACTCGCAGCGCTGSGTGGCGCGAC 351  
 Qy 340 TGGGGGCGATACGGGGGTTTGTATGATCCGTATGGCTTGGCACTCCGCTGGCACTACCGT 399  
 Db 352 TACGGCCGATACGGGCGGCTGTATTCGCGATGGCGTGGCACTTCGCGGCACTACCGC 411  
 Qy 400 ATTGCTGATGCGCGTGGCGGCTGTGACGAGCGAGCAGCGGTTCGCGGCTGCACTGTC 459  
 Db 472 TGGCGCGCAACACCCACCTTGGCAAGCGCGCGGCTGCTGTGGCGGTCAAGAAG 531  
 Qy 520 TACGCGCAACAAATCACTGGCGACAGCTGATGATTCGTGGCGACCGTGGCTATGAC 579  
 Db 532 TACGCGCAAGAACTCTCATGGCGGAGCTGATGTTTTCGCGGCACTGCGGCTTGAA 591  
 Qy 580 TCGATGGGCTTACCTGCTTACGGCTTCTCTTTTCGGCGCGGTGCGATTTTGGGAACCG 639  
 Db 592 TCGATGGGCTTCAAGACGTTGCGGTTTCGGGCTTCGGCGGCTCGACAGTGGAGCGCG 649  
 Qy 640 AAGATGATCTACAGGCTGACGAAAGAGTGGCTGGCAACCTTCTGACCAAGCTGACCG 699  
 Db 650 -ATGAGGCTTATTGGGGCAAGGACCACTGGCTGCG-----GCGATGAGCGTTACAGC 702  
 Qy 700 GACGTGAACAGCAGAGACCATGGAACACCGCTGGCGGCTGTCCAAATGGGTCTGATC 759  
 Db 703 GGTAAAGGGATC-----TGGAGAACCGCTGGCGGCTGCGATGGGCTCATC 753  
 Qy 760 TATGTGACCGGGAAGTGTTAACGCGCACCTGATCGCTGAGAACGCGCAGCAGGTA 819  
 Db 754 TACGTGAACCGGAGGGCGGCAACCGGCAACCCGAGCCCATGCGCGGCGGCGTGCACAT 813  
 Qy 820 CTTGAAACCTTCGCGCCGTATGGCGATGAACGACAAAACCGAGCGCTTCACAGCTGCG 879  
 Db 814 CCGAGAGCTTTCGGCGCATGCGCATGACGACGTGGAACGCGCGGCTGATGTCGCG 873  
 Qy 880 GCGCACACCGCTCGGATTTGTCAGCGTAATGGCTATGCTGCTGCTAGCCCTGACCA 939  
 Db 874 GGTCAACATTTTCGCTTAAGACCCATGCGCGCGCGCGCGCGGATCTGTGGCGCCG 933  
 Qy 940 AAAGCCTCTGACGTTGAAAACAGGCGCTTAGGTTGGGGCAACCCCAACATGACGGGCA 999  
 Db 934 GAGGCTGCTCGCTGGAGCAGTGGGCTTGGGCTTGAAGAGCTCTGTATGGCACCGGAC 993  
 Qy 1000 GCAACGACCGCTGACCTCGGATTCGGAAGTGTCTTGGACCAACCAACCGCAAGAA 1059  
 Db 994 GGTAAAGACGCGATCACCAGCGGCATCGAGGTGTATGACGAAACCCCGCAAGAA 1053  
 Qy 1050 GATATGGGCTATTTCGACCTCTGTTTCGGGTACAAATTTGGGAACCTGAAAGAAGTCTCC 1119  
 Db 1054 GACACAGTTTCTCGAGATCTGTGACGGCTAGCGGTGGGAGGCTGAGGAGCGCTGCT 1113  
 Qy 1120 GGTGCGCCACATTTGGGACCGGATTTGACATCAAAAGGAAGAACGCGGTGAGCGCA 1179  
 Db 1114 GCGGCTTTGGCAATACACGCCCAAGAGACGGCGCGGTGGCGCACCACTATCGCGAC 1173  
 Qy 1180 GACCCCTCTATTTCGCGCAACCGCGATGACCGATTCGGGATATGGGCAAGAGTAAT 1239  
 Db 1174 GCGGCGCC---AGGCGGCTCCCGGAGATCTGGCCACTGACCTGCTGCTGCGGGTGGAT 1230

QY 1240 CGSACCTATGCGCGCTATCTCGCGAAATATTCATGCGCGGATCTCTGATGACTTCTCAAGAAACT 1299  
 Db 1231 CCGATCTATGAGCGGATACACGCTGCGTGGCTGGACACCCGAGGAATGGCGCCGAG 1290  
 QY 1300 TTCCGGAAGAGCGGTGTTAAAGCTGACGACGCTGAGCGCCGGAATATCACTGTACATC 1359  
 Db 1291 TTTCGCCAAGCGCTGATACAGCTGATCCACAGAGATGGGTCCGGTGCAGATACCTT 1350  
 QY 1360 GCGCGAGATGTCGCGGAGAGCTATGTCGCGAGACCGGATTCGCGAGGTACAC 1419  
 Db 1351 GGGCGCTGGTCCCGAAGCAGCCCTGCTGGCAGGATCGGTCCGCTCAGCCAC 1410  
 QY 1420 GACTACTCGGAAGAAGTGC-----TCACGACAAATATTCACAAAGTGGCTG 1467  
 Db 1411 GACCTCGTGGGGAAGCGAGATATCCAGCCTTAGAGACCCAGATCGGGGATCGGATG 1470  
 QY 1468 AGCATTTAGTGTATGAGTGTGCGACGCTTGGACAGTGGCCGCTACTATGCGGTTCCGAT 1527  
 Db 1471 ACTGTCTCACAGTAGTTTCAGCCGATATGGCGGCGGCTGCTGCTGCTGCTAGCGAC 1530  
 QY 1528 ATGCGGCGCGGTGTACGCTGGTGGCCGATTCGTTGGCCCGACAGAGAGTGGCAGGCG 1587  
 Db 1519 AAGCGGCGCGGCGGCGAGCGTGTGCTGCGCTGCGTCCAGCCAAAGTGGGAGGCTC 1590  
 QY 1588 AAGCGCGCGGA-----CGGCTGGCGAAGTGTGAGGCTGTACGAGAGAGCT----- 1636  
 Db 1591 AAGGACCCGAGGGGATGTGCGCAAGTCAATTCGACCTCGGAAGATCCAGGATCA 1650  
 QY 1637 -----CTGCGGACACGGGCGTATGCTACGCGGAGCTGATCTTCTGGCG 1680  
 Db 1651 TTCMACTCCGCGGCGCGGGGACATCAAGTGTCTTCGCGACCTGCTGTGCTGCTG 1710  
 QY 1681 GGTAGGCTTAGCATCGAGAGAGCGGCGGAAAGCGAGGTTAGCATGTGCGGCTTCCCTTC 1740  
 Db 1711 GCTGTGCGCCATAGAGAAGACGAAAGCGGCTGGCCCAACATACAGTGCCTTC 1770  
 QY 1741 CTGAAAGCGGTGGCGATGCGACCGCCGAGATGACGAGCAGACTCTTGCACCGCTG 1800  
 Db 1771 ACCCGGGCGCGCGATGCTGCGAGACGACGAGTGTCTTCGCGACCTGCTGTGCTGCTG 1830  
 QY 1801 GAGCGCTGCGCGATGGCTTCCGCAACTGTGCGAGACGAGAGATGTGCTGAGCGCGGA 1860  
 Db 1831 GAGCCCCAAGCAGATGGCTTCCGAACTACTCGGAAAGGCAACCGCTTCCCGGCGGAG 1890  
 QY 1861 GAGATCTGCTGGATGTCGCGAGCTGATGGGCTTACCGCGCCGGAATGACGCTGCTG 1920  
 Db 1891 TACATCTGCTTCGCAAGCGGACGCTGTACGCTAGTGCCTTCGAGTGAAGCTGCTG 1950  
 QY 1921 CTGGCGGATATGCGCTATGGGCAACCACTATGTGTGCGCAACAGCGCGTATTCCAC 1980  
 Db 1951 GTAGTGGCGCTGCGCGCTCTGCGGCAAACTACAGGCGCTTACGCGTGGGCGTGTTCAC 2010  
 QY 1981 GATTTGAGGCCAGTTTGACCAACGACTTTTGTGAACTCTGACCATGTCCGAGACAGC 2040  
 Db 2011 GAGGCTTACGCTATGACCAACGCTTCTGTGAAACCTCTCTGACATGGGTATCC 2070  
 QY 2041 GAGGAGC---CGTAGTGAAGACCCCTACGAAATCCGAGCCGACGAGCGTGGCGTG 2090  
 Db 2071 TGGGAGCCCTCGCCAGCAGATACGGGAGCTTACCGAGGCAAGGATGGCAATGGG 2130  
 QY 2098 AAGTGAAGCGCTCTCGGCTGAGATCTGGTATTGTGTTCCMACTCGCTTCTGCTCTTAC 2157  
 Db 2131 AAGTGAAGCGGAGCGCTGAGGAGCTTGTGTTGGTTCMACTCGGAGTGTGGGCGCTT 2190  
 QY 2158 GCAGAAGGTGTACGCCGACGAGATACCGCGAGAGATTTCTCAGAGACTTCTGCGCGCG 2217  
 Db 2191 GTGAGGTCATTGTGCGCGCGATGATACCGCAGCGCAAGGTTCTGCAGGACTCTGCTGCGC 2250  
 QY 2218 TGAACCAAGTGTGAAGCGGAGCCTGTTTGAGCT 2252  
 Db 2251 TGGGACAGGCTGTGATCTGACAGAGCTTCAGCT 2285

## RESULT 9

US-09-103-840A-2/c  
 : PATENT NO. 6254128  
 : PUBLICATION US/09103840A

: GENERAL INFORMATION:  
 : APPLICANT: FLEISCHMAN, Robert D.  
 : APPLICANT: FRASER, John R.  
 : APPLICANT: WHITE, Owen R.  
 : TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 : TITLE OF PUBLICATION: US/09103840A  
 : FILE REFERENCE: 24366-20007.00  
 : CURRENT APPLICATION NUMBER: US/09/103.840A  
 : CURRENT FILING DATE: 1998-06-24  
 : NUMBER OF SEQ ID NOS: 2  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 2 4403765  
 : TYPE: DNA  
 : ORGANISM: Mycobacterium tuberculosis  
 : FEATURE:  
 : OTHER INFORMATION: CDC 1551  
 : OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 : OTHER INFORMATION: represent a, t, c or g  
 : US-09-103-840A-2

Query Match 27.3%; Score 618.6; DB 3; Length 4403765;  
 Best Local Similarity 58.7%; Pred. no. 1,8e-162;  
 Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;

QY 160 AACAAAGATTGTTGGCCCGAAGGTTGAACCTGTGATATTTTGCATCAGCAAGATCGCAAA 219  
 Db 2153267 AACCGAGACTGTGTGGCCCAACCGGCTCACTCTGAAGTACTGCACAAACCGCGCGTC 2153208  
 QY 220 TCAGACCGGATGATCCGAGTTTTCACCTACCGTGAAGACGACGACGCTGATTTGGAG 279  
 Db 2153207 GCTGACCCGATGGGTGCGGCTTTCGACTATGCCGGGAGTGGCGACCATGACGTTGAC 2153148  
 QY 280 GCGCTGAAGAAAGATGTCACGGGTTGATGACCGATAGCCAAAGAGTGTGGCCCGCTGAC 339  
 Db 2153147 GCGCTGCGCGGACATCGGAGAGATGACCCCTCGCACCGCTGTGGCGGCTTTACAC 2153088  
 QY 340 TGGCGGCACTACGCGGTTTGTATGATCCGTTATGCTGTGGCAGCTCCGCTGCGCACTACCT 399  
 Db 2153087 TACGGCACTACGCGGCGCTTATATCCGAGTGGCTGCGCAGCTGCGCAGCTACCGC 2153028  
 QY 400 ATTGCTGATGGCGCTGGGGCGGTGTACCGGAAGCGCAGCGCTTTCACGCTCACTCC 459  
 Db 2153027 ATCCACGACCGCGCGCGCGCGCGGCGGCGCATGAGCGGTTCGGCGGCTTTACAC 2152968  
 QY 460 TGGCGGCAACAGTCAGCTGAGTAAGGCGGCGCTGCTGTGGCGGATCAAGAAGAG 519  
 Db 2152967 TGGCGGCAACAGCAGCTTGGACAGGCGCGCGGCTGCTGTGGCGGATCAAGAAGAG 2152908  
 QY 520 TACGGCAACAAATCATGCGGCGAGCTGTATCTGCGCTGCGCACGCTGCTGATGAG 579  
 Db 2152908 TACGGCAACAAATCATGCGGCGAGCTGTATCTGCGCTGCGCACGCTGCTGATGAG 2152848  
 QY 580 TCAATGGGCTTACTGCTTACGCGTTCTGTTCGCGCGCTCGATATTGGGACACCCGAA 639  
 Db 2152847 TCGATGGGCTTCAAGACGTTGCGGCTTGGCGGCTGCGCGGCTGCGACGCTGGAGCCGC 2152790  
 QY 640 AAGATATCTACGTGGGTGAGAAAAGAGTGTGGAGACTCTGTGACGAACGCTGACGCG 699  
 Db 2152789 -ATGAGTCTATTGGCGAGAGAACCCCACTGGCTG-----CCGTAGAGGTTTACAC 2152737  
 QY 700 GAGCTGAACAACCCAGACACCATTCGAAAAACCGCTGCGGCTGTCCAAATGGGTCTGATC 759  
 Db 2152736 GTTAGCGGGGATC-----TGGAGAACCCCGCTGCGCGGCTGCAGATGGGCGCTATC 2152686  
 QY 760 TATGTGAACCCCGAAGGTTTATGGGCAACGCTGTATCCGCTGTGAGAACCGCAACGAGTGA 819

[illegible]

QY	1861	GAGATGCTGCTGGATGTGGCGACAGCTATGCGGCTTACACGGCCGGAAGATGACGTCGTG	1920
Db	2151548	TACATCTGCTCTGCACAGCGGACCTCTTACGCTCAGTGGCCCTCTGATATGAGGTGCTG	2151489
QY	1921	CTGGCGGGATATGCGCGTACTGGGGACCACACTATGTGGCGACCAACACGAGCGCTATTACCG	1980
Db	2151488	CTAGGTGGCCCTGCGCGTCTCTGGCGGCAAACTACAGAGCGCTTACCGCGTGGCGGTGTTCAACC	2151429
QY	1981	GATGTGGAAGCGGCAAGTTGACCAAMCAGCACTTTTGTGTGAACCTGACCAATATGGGGACAGC	2040
Db	2151428	GAGGCTTCCGCTCACTGACCAACACGCTTCTGTGGAACCTCTCGACATGGATATGATACACC	2151369
QY	2041	TGGAAGC--CGGTAGTAGCAACGCTTACGAATACGCGGACCGCAACGACCGGTGGCGGTG	2097
Db	2151368	TGGGAGCCCTCCGACGACAGATGACGGGACCTACCAAGGSGCANGATGCGAGTGGCAAGGTG	2151309
QY	2098	AACTGGACCCGCTCGGGGGTGGATCTGTTATGTTGGTCTCAACTGGCTACTGGGCTTTAC	2157
Db	2151308	AACTGGACCCGCGCGCTGGTGGACACTGGTCTTGGGTGTCACACTGGGATTTGGCGGGCGT	2151249
QY	2158	CGAGAGGTCTAGCGCCACGACGACATGAAGCGGAGAGTGTGTCAGAGACTCTGCGCGGCGC	2217
Db	2151248	GTGAGGTCTATGTGGCGCGATGACSCGACGCGGAAGTTCTGCGAGGACTTCTGCTGCTGCC	2151189
QY	2218	TGGACCAAGTGAATGAACGCGGACCGTTCGAGGT	2252
Db	2151188	TGGGACAAAGGTGATGAACCTCGACAGGTTCGAGGT	2151154

RESULT 10

US-09-103-840A-1/c

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: FRANK, John

; TITLE OF INVENTION: TUBERCULOSIS

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103-840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 4411529

; ORGANISM: Mycobacterium tuberculosis

; OTHER INFORMATION: H37Rv

US-09-103-840A-1

Query Match 27.3%; Score 618.6; DB 3; Length 4411529;

Best Local Similarity 58.1%; Pred. No. 1.8e-162;

Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;

QY 160 AACNAGATGTCGGCGGAGAGGTTGACTCGATATTTCGATCCAGACGATCGGAA 219

Db 2156007 AACCAAGACTGTGTGGCCCAACCGGCTCAATCGAAGGTCTGACCAAAACCGGCGCTC 2155948

QY 220 TCAGACCCGATGATCGGATTTTCAACTACGCTGAAGAAGTACGAAAGCTGATTTTCGAC 279

Db 2155947 GCTGACCCGATGGGTGGGCGTTCGACATATCGCGCGGAGGTGCGGCACATCGACGTTGAC 2155888

QY - 280 GCGCTGAAGAAAGATGTCACCGCTTGATGACGATAGCCAAAGATGGTGGCCCGCTGAC 339

Db 2155887 GCGCTGACCGGACATCGGAAGATGTGACNCTGACGCCGCTGGTGGCCCGCGAC 2155628

QY 340 TACGGGCACTACGCGGCGCGCTTTTATCGGATGGGTGTGGCTGGCATCCGCTGGCACTACCGT 399

Db 2155827 TACGGGCACTACGCGGCGCGCTTTTATCGGATGGGTGTGGCTGGCATCCGCTGGCACTACCGT 2155768

QY	400	ATTGCTGANTGGCCGCTGGGGGGGGGGTGATACCGGAAGCAGACGCGCTTGGACCGCGCTCAACTCC	459
DB	2155767	ATCCAGCAAGCGCGCGGGCGCGCGCGCGCGGCGGCGATACAGCGGTTCCGCGCGCTTACAGC	2155708
QY	460	TGGCGCGCAACGCTCAGCCTTGGATAAAGGGGGCGCTCTGCTGTGGCGCATCAAGAAGAG	519
DB	2155707	TGGCGCGCAACCGCAGCTTGGCGAGGCGGCGCGCGCTGTGCGCGGTCAACGAG	2155648
QY	520	TAGCGCAACAAATAGCTTGGCGAGCGCTGATCTGCTGTGCGCGCGGTGCTTACAG	579
DB	2155647	TAGCGCAAGAGCTCTCAATGGGCGGCGCATGATGTGTTTTGGCGGCGACATGGCGCCTGGAA	2155588
QY	580	TCCATGGGCTTACCTGCTTATAGGCTCTCTTTCGGCGCGGCGTGATATTGGGAACCCGAA	639
DB	2155587	TGCTGGGCGCTTCAGAGCTGTGGGCTTCGCTTGGCGCGGCTTGCAAGTGGAGCGCG--	2155530
QY	640	AAAGATATCTACTCTGGGCTGACAGAAAGAGTGGCTGGCACCTTCTGAGCAACGCTTACGGC	699
DB	2155529	ATGAGGCTTATTTGGGCGCAGAGAGCGACCTTGGCTCG-----GCGATGAGCGTTTACAGC	2155477
QY	700	GAGTGTGACACCGACGACCTTATGGAAACCCCGCTTGGCGGCTTGCTCAATATGGCTTGATC	759
DB	2155476	GGCTAGCGGGGATG-----TGGAGAGCGGCTTGGCGCGGCTTGATATGGGCTTGATC	2155426
QY	760	TATGTGAACCCGGAAGTGTTTAAAGCGCACACCTTGATCTGAGTACCGCAGCAGCATGA	819
DB	2155425	TAGCTGTAACTGGGAGGGCGGAGCGCAACCGACCGACCCATGCGCGCGCGCTGCACATT	2155366
QY	820	CTTGAACCTCTCGCCCGTATGGCGATGACAGCAAAACCGGAGCCCTTCACAGCTGGC	879
DB	2155365	CGGAGAGCTTTCGGCGCATTGCGCATGACGCGGCTGCGAGAGCGCGCTTATCTCGCGC	2155306
QY	880	GGCCACACGCTCGGTAACTGTACAGGTAAATGGCAATGCTCTGCTGTAGCGCCTTCAACCA	939
DB	2155305	GCTCACTATTCGGTTAAGACCCATGATGGCGCGCGCGCGCGATGCTGGTGGCCCGCGAAACC	2155246
QY	940	AAAGCTCTGAGCTGTGAAACCGAGTGTAGGTTGGGGGACCCCAATGTCAGGCGCAG	999
DB	2155245	GAGGCTGCTCGCCTGGACAGATGGGCTTGGAGAGCTGATATGGCACCGGACACC	2155186
QY	1000	GCAAGCAACCGCTGACTCGGGGTATGAGAGTGGCTTGGAGACCAACCCACAGAAATTC	1059
DB	2155185	GGTAGAGAGCGATACACCGCGGATCGAGTGGTATGAGAGCAACCCCGACGAAATGG	2155126
QY	1060	GATATGGGCTATTTCGACCTGCTGCGTACATATTGGTATGAGTGTGAAGAGTGTCTGCC	1119
DB	2155125	GACACAGTTTTCCTCCAGATCTGTACGGCTACGAGTGGAGCTGACGAGACCGCTGCT	2155066
QY	1120	GGTGCGCCACCATTTGGGACCGATGTGACATCAAAAGAAACCAAGCGGTTTGAGCGCAG	1179
DB	2155065	GGGCGCTTGGCATACACCGCGAAGGAGCGCGCGTTCGGCGACATCCCGAGCCGCTTC	2155006
QY	1180	GAGGCTCTATTCGCGACACGATACATCGGATCGGATATGGGATGAGGATTAAT	1239
DB	2155005	GGCGGGC-----AGSGGGCTTCGCGAGATGCTGGCGCATGACCTTGCCTGGGTGGAT	2154949
QY	1240	CCGACATATCGCGCTATCTCGCAAAATTCATTTGGCGGATCTCTGAGTACTTCAAGAAACT	1299
DB	2154948	CGGATCTATGAGCGGATACCGGCTGCTGGCTGGAAACCCCGAGAAATTTGGCGACAG	2154889
QY	1300	TTGCGGAGGCGTGGTTTCAGATGCGGAGCGAGTGTGAGCTTGGCGCGGAATTCAGTTTACATC	1359
DB	2154888	TTTGGCAAGGCTGCTCAAGCTGATCCAGCGAGATGSGTCCGCTGTGGGATACCTT	2154829
QY	1360	GGCCCGGAGTGGCGCGAGAAAGCTGATTTTGGCAACCGCATATTCGGCGAGTTAAACAC	1419
DB	2154828	GGGCGGCTGTCTCCCAAGCAGACCTGCTGTGTGGAGAGTCTGGGTCTCTCGGTCAAGCAC	2154769
QY	1420	GACTCTGCGGAGAGTGG-----TCAAGCAAAATTCACAAAGTGGGCGCTG	1479
DB	2154768	GACTCTGCGGCGAGCGGAGATTTGCAAGCTTGGACGATTCAGGCGGATCGAGGATTCG	2154709
QY	1468	GACCTATTGATGATGCTCCACCGGCTGTGGACAGTGGCCGCTACTTATTCGGGCTTCGAT	1527

Db 2154708 AGCTGCTCCACAGCTAGTGTTCGACCCGATATGGCGGGGGGGCTGCTGTTTCCGTGGTAGCCGAC 2154649  
 QY 1528 ATGCGCGCGCGGTGTCATACAGGTGTCGCCGATTCGGTTTGGCCCCACAGAACGAGTGGCAGGGC 1587  
 Db 2154648 AAGCGCGCGCGCGGACGAGTGGTTCGCTTCGCGCTCTGAGGACCAAGTGGGTGGGGAGTGC 2154589  
 QY 1598 AAGGAGCGCGA---CGCGCTGGCGAAGAGTGTGAGCGTGTACGACAGACATCT-----1636  
 Db 2154588 AACGACCCCGACGGGGATCTGGCGAAGGTCTATTCGCACCTGGAGAGATGACGAGTCA 2154529  
 QY 1637 -----CTGCGGCACACCGCGCTAGCATATCGCGGACGATGATCTTCCGACCGGTG 1680  
 Db 2154528 TTCATCTCCGCGGCGCCCGGGGAACATCAAGTGTCTTTCGCGCATCTTGGCTTCGCT 2154459  
 QY 1661 GTTAGCGCTAGGCGATGAGAACGCCCGCGAGACAGCAGCATGTTATCGATGCGCGTCCCTTCG 1740  
 Db 2154468 GCCTGTGCGGCCATAGAGAAGCGACAAAGCGCGCTGGCGACAAACATCAGCGTGGCTTTC 2154409  
 QY 1741 CTGAAGAGCGGTGGCGATGGCGACCGCGAGATGACGAGCGAGACTCTTCCGACCGGTG 1800  
 Db 2154408 ACCCGGGCGCCCGAGTGTGTGGAGAGCAACCGACGTGAATCTTTCGCTGCTG 2154349  
 QY 1801 GAGCGCGTGGCGGATGCTTCGCGAATCTGGGCAAGAGAGATGATGGTGAAGCGGNA 1860  
 Db 2154348 GAGCGCAAGCGAGATGGCTTCGGAACATCTCTGGGAAGGCGCAACCCGTTCGCGCGCGAG 2154289  
 QY 1861 GAGATGCTGTGGATGTGGCGAGCTATGGGCTTACCGCGCCCGGAATACGCTGCTG 1920  
 Db 2154288 TACNCTGTCTGCGCATGCTGGCGAGCTCTTACGCTCAGTGGCGCTCGATGACGTGCTG 2154229  
 QY 1921 CTGCGCGGTATGCGCATGCTGGGCAACCATCTTACGCTCAGTGGCGCACTACAGCGCGTATCCAG 1980  
 Db 2154228 GTAGTGGCGTCTGCGGCTCTCTGGCGCAAACTACAGCGCTTACCGCTGGCGGTGTCAAC 2154159  
 QY 1961 GATTGTGAGGCCAGTGTGACCAACGACTTTTTTGTGAACTGACCGATATGGGGAACAGC 2040  
 Db 2154158 GAGCGCTCCGATCTATGACCAACGACTTCTGTGAACTGTCTCGACTGGGTATCCAC 2154109  
 QY 2041 TGGAGC---CGGTAGGTAGCACGCTTACGAAATCGGCGACGCGAAGCCGTCGGTG 2097  
 Db 2154108 TGGAGGCGCTTCGCCACAGATGACGCGGACTTACAGGCGAGATGGCATGGCAGGTG 2154049  
 QY 2098 AAGTGAACCGCTTCGCGGTGATGTGGTATGGTTCACACTGCTACTGCGGCTCTTAC 2157  
 Db 2154048 AAGTGAAGCGTACCGCGTGGAGTGTGTGGTGTGCTGCGTGATGTTCGGGGCGCTT 2153989  
 QY 2158 GCAGAGGTGTAGCGCCAGGACAGTACAGCGGCAAGTGTGTGAGAGATTTCTGTCGCGCGC 2217  
 Db 2153988 CTGAGGTGTATGGCGCGGATGAGCGCCGACGAGTGTGGTGCAGGACTTCTGCTGCC 2153929  
 QY 2218 TGGACAAAGTGTGAGCGCGACCGCGTTPCCAGCT 2252  
 Db 2153928 TGGGACAGGTGTGACCTTCAGAGGTTTCAGCT 2153894

RESULT 11  
 US-08-418-782-1  
 : Sequence 1, Application US/08418782  
 : Patent No. 5658733  
 : GENERAL INFORMATION:  
 : INVENTOR: K. Kozhili, Franklin R.  
 : APPLICANT: K. James R.  
 : APPLICANT: K. James R.  
 : TITLE OF INVENTION: Detection of Isoniazid Resistant Strains  
 : NUMBER OF SEQUENCES: 22  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Schwegman, Lundberg & Woessner  
 : STREET: 3500 105 Center  
 : STATE: MN  
 : CITY: Minneapolis  
 : COUNTRY: USA

RESULT 11  
US-08-418-782-1  
: Sequence 1, Application US/08418782  
: Patent No. 5658733  
: GENERAL INFORMATION:  
: APPLICANT: Cockerill, Franklin R.  
: APPLICANT: Kline, Bruce C.  
: APPLICANT: Uhl, James R.  
: TITLE OF INVENTION: Detection of Isoniazid Resistant Strains  
: TITLE OF INVENTION: of M. Tuberculosis  
: NUMBER OF SEQUENCES: 22  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Schwegman, Lundberg & Woessner  
: ADDRESS: 3500 11th Center  
: CITY: Minneapolis  
: STATE: MN  
: COUNTRY: USA







```

Db 1114 GCGGGGCC---AGGGCGTCCCGCAGCATCTGGGCACCTGACCTCTCGCTCGGGGTGGAT 1170
QY 1240 CCGACCTATCGGCCTACTTGGGAAATATCATGGCGGATCTCTGACTACTCAAGAAACT 1299
Db 1171 CCGATCTATGAGCGCATCAGCGTCTGCTGGCAGACACCCCGAGGAATTGGCCGACGAG 1230
QY 1300 TTGCGAGCGGTGGTCTCAAGCTGACGCACGCTGAGCCGCCGCAATCACGTTACATC 1359
Db 1172 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1231 TTGCGAGCGGTGGTCTCAAGCTGACGCACGCTGAGCCGCCGCAATCACGTTACATC 1290
Db 1173 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1360 GCGCGCGAGTGGCGGACAGCATCTTGGCAGACCGGATTTCCGCGAGGTACACCC 1419
Db 1174 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1291 GGGCGCGTCTGCGCCACAGCACGCTTGTGTGCGAGSATCTCCGGTCCCTCGCGTACGCCAC 1350
Db 1175 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1420 GACTACTGCGAAGMTGG-----TCAGAGAAATTTGCACAAAGTGCGCTG 1467
Db 1176 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1351 GACCTAGTGGAGTGGCTCTGCACATTCGTCGAGCTTTAGAGCCAGATCCGGGATG 1410
Db 1177 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1468 GACATTAGTGGAGTGGCTCTGCACATTCGTCGAGCTTTAGAGCCAGATCCGGGATG 1410
Db 1178 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1411 ACTGTCTTCACAGCTATGTTGACGCGATTTGGCGCGCGGCGCTGCTGCTTGGTACGCGAC 1470
Db 1179 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1528 ATGCGCGGGGTGCTACAGGTGCGCGGCTTTCGTTGTGTCGTCGCGCGGCGGCGGCG 1587
Db 1180 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1471 AAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1530
Db 1181 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1588 AACGACGCGCGA---GCGCGCTGGGAAAGTGTGAGGTCTACAGCAGATCT----- 1636
Db 1182 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1531 AAGCAGCCGACGCGGGATCTTGGCAAGGTCTTGGCACCTTGGAGAGATCCAGAGTCA 1590
Db 1183 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1637 -----CTGCGACACCGCGCTAGCATCGCGGAGGTGATCTTCTTGGGCC 1680
Db 1184 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1591 TTCACCTCGCGCGCGCGGAGATCAAGGTCTCTCGCGCGCTCTGCTGCTCGGT 1650
Db 1185 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1681 GGTAGCTTAGCATGAGAAAGCCGCGGAAAGCAGAGTACGATGTGCGGTTCCTTC 1740
Db 1186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1651 GCGTGTGCGCGCATAGAAAGACCAAGCGCGCTGCGCACATCATCAGGTGCGCTTC 1710
Db 1187 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1741 CTGAAAGCGCGTGGGATGTCAGCGCGCGGAGATGACGAGCAGCTCTTTCGACCGGTG 1800
Db 1188 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1711 ACCCGCGCGCGCGCGTTCGCGAGGACAAACGCGGTGATCTTTCGCTGCTG 1770
Db 1189 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1801 GAGCGCTTGGCGGATGCGCTTCCCACTTGGCAGAAAGAGATGTGTGGTGAAGCGGAA 1860
Db 1190 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1771 GAGCGCTTGGCGGATGCGCTTCCCACTTGGCAGAAAGAGATGTGTGGTGAAGCGGAA 1830
Db 1191 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1861 GAGCGCTTGGCGGATGCGCTTCCCACTTGGCAGAAAGAGATGTGTGGTGAAGCGGAA 1920
Db 1192 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1831 TACATCTCTCGACAGCGCATCTGCTTACGCTTANGTGGCCCTGAGATGAGTGTGCT 1890
Db 1193 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1921 CTGGGCGGTATGGCGGTACTGGGCGCAATATGTGTGCACCAACAGGGGTATCTTACCC 1980
Db 1194 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1891 GTAGTGGCGGTGCGCGGTCTGCG---GCAACTACAGCGCTTACCGCTGGGCGTGTTCACC 1949
Db 1195 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1981 GATGTGAGAGCCNATGACCAACGACTTTTGTGAGCTGTACCGATGTGGGAGACG 2040
Db 1196 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 1950 GAGCGCTCGGATGCTACTGCAACGACTCTTGTGAGCTCTGCTCGATGGGTATCTACC 2009
Db 1197 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 2011 TGGAGCG---CGGTAGTGTAGCAGCGCTTACGAAATCCGCGACCGCAAGACGCGTGG 2097
Db 1198 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 2040 TGGAGCGCGCTCGCGGTGAGTGGCGGACCTACCGAGGCAAGGATGGCATGTGGCAAGGTG 2069
Db 1199 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 2098 AAGTGTGAGCGCTCGCGGTGAGTGGCGGACCTACCGAGGCAAGGATGGCATGTGGCAAGGTG 2157
Db 1200 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 2070 AAGTGTGAGCGCGCGGTGAGTGGCGGACCTACCGAGGCAAGGATGGCATGTGGCAAGGTG 2129
Db 1201 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 2158 GCAGAGCTATGACGCCAGAGCATATACGGGCGAGAGTTCGTACAGAGATCTGTGGCGGCC 2217
Db 1202 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 2130 GTCGAGGTCTATGGCGCGGATGAGCGGACCGCGAGGTGTGTCAGGACTTGTGCGTCC 2189
Db 1203 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 2218 TGGCAAGAGGTATGAGCGCGCGCGGTGAGCT 2252
Db 1204 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1310
QY 2190 TGGGACAGAGTGTGATGCTCGACAGGTTCGAGCT 2244

```

## RESULT 13

```

US-08-852-219-1
: Sequence 1. Application US/08852219
: Patent No. 5922575
: GENERAL INFORMATION:
: APPLICANT: Cockerill, Franklin R.
: INVENTOR: Kline, Bruce C.
: TITLE OF INVENTION: Method of Isoniazid Resistant Strains
: TITLE OF INVENTION: of M. Tuberculosis
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Mueling, Raasch & Gebhardt, P.A.
: STREET: 119 No. 5922575th Fourth Street, Ste. 203
: CITY: Minneapolis
: STATE: MN
: COUNTRY: USA
: ZIP: 55401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT INVENTION DATE: US/08/852, 219
: FILING DATE: 07-May-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Sandberg, Victoria A.
: REGISTRATION NUMBER: 41,287
: REFERENCE/JACKET NUMBER: 230.00010130
: TELEPHONE: 612-305-1226
: TELEPHONE: 612-305-1226
: TELEFAX: 612-305-1228
: INFORMATION FOR SEQ ID NO: 1:
: LENGTH: 2235 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
US-08-852-219-1

```

Query Match 26.8%; Score 606.6; DB 2; Length 2235:  
 Best Local Similarity 58.6%; Pred. No. 1.1e-160;

Matches 125; Conservative 0; Mismatches 819; Indels 64; Gaps 9;

```

QY 160 ACAAGATTTGGTGGCGGAGGCTTGAACCTGGATTTTGGCATGAGAGATCGCAAA 219
Db 112 AACCAAGCTTGGTGGCGGACCAACCGGCTCAACTGAAGTACTTGCACAAACCCGCGCTC 171
QY 220 TCAGACCCGATGGATCCGGATTTTCACTACCTGAGAAAGTACCAAGCTGGATTTTCGAC 279
Db 172 TCGACCCGATGGGTTGGCGGCTTGCATTTGCCGGGAGTCCGCCATCTGACGCTGC 251
QY 280 GCGCTTGAAGAAAGTGTACACGCGTGTGATGACGATAGCAGATGAGGTGGCGCGTCAC 339
Db 232 GCGCTTGAAGAAAGTGTACACGCGTGTGATGACGATAGCAGATGAGGTGGCGCGTCAC 291
QY 340 TGGGCGGATTCAGCGCGTGTGATATCCGATTCGCTATGGCTTGGCATTCGCTTGGCACTTACCGT 399
Db 292 TACGGGCTTACGCGCGCTGTTTATTCGGTGGCGGACCTTCCGGCACTTACCGT 351
QY 400 ATTGCTGATGGCGTGGCGGCGGTGTACCGGAGCGACGCTTTGCACCGCTCAACTCC 459
Db 352 ATCCACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 411
QY 460 TGGCGGACCACTGTACGCTTGATTAAGCGCGCGCGCTTCGCTGTGGCGGATCAAGAAGAG 519
Db 412 TGGCGGACCACTGTACGCTTGATTAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 471

```



```

: FILLING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Messner, Warren D.
: REGISTRATION NUMBER: 30,440
: REFERENCE/DOCKET NUMBER: 150,141US1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-339-0331
: FAX: 612-339-0331
: INFORMATION FOR SEQ ID NO. 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2221 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MODIFICATION: PE: DNA
US-08-418-782-2

Query Match      25.68; Score 578.2; DB 1; Length 2221;
Best Local Similarity 57.88; Pred. No. 1e-152;
Matches 1229; Conservative 0; Mismatches 838; Indels 58; Gaps 9;

QY 160 ACAMAGATTTGGTGGCGAAGAGGTGTGAACCTGATATTTTGGATCAGCAGAGTAGCGAA 219
DB 112 AACGACAGCTGGTGGCCCAACCGCGCTCATCTGAGAGTACTGCACAAACCGCGCGCTC 171
QY 220 TCAGACCGATGATGATCCGGATTTTCAACTACCGCTGAGAGAGTACGCAAGCTGATTTGCAG 279
DB 172 GCTGACCGGATGGGTGGCGGCTTTCGACTATCGCGGGAGTGGCGGACCAAGTGCATTTGAC 231
QY 280 GCGTCGAAAGAGATGTCAAGCGGTTGATGACGATAGCCCAAGAGTGTGGTCCGCTGTAC 339
DB 232 GCGCTGACGCGGACATCGAGAGAGATGATGACCACTCGGACCCGTGTGGCGCGCGAC 291
QY 340 TGGGGGCACTACGCGCGGTTTGATGATCCGATATGGCTTGGCACTCCGCTGCGACCTACCGT 399
DB 292 TACGGCAGCTACGCGGCGCGTGTATTCGCGATGGGTGGTGGCAGCTCGCGCGACCTACCGC 351
QY 400 ATTCGCTGATGCTGGGGGGGGTGGTATCGGAAGCGAGCGGTTTGACGCTGCACTCC 459
DB 352 ATCCAGCAGCGCGCGCGCGCGGGGCGGATGACGAGCGGTTCGCGCGCACTTACACG 411
QY 460 TGGCGGACACAGCTCAGCTTGGATTAAGAGGGCGCGTCTGCTGGCGCGATTCAGAAAGA 519
DB 412 TGGCGGACCAAAATCAGCTTGGACAAAGGGCGCGCGCTGTGGCGCGGTCAAGAAAG 471
QY 520 TACGGCGCAACAGCTGCGGCAAGCTGATGATTCGGCTGCGA--CGTGGCTATTG 577
DB 472 TACGGCAAGAAGCTCTCATGGCGGACCTGATGTTTTCGCGGCAACCGCTGCCCTGCG 531
QY 578 AGTCATATGGGCTTACCTGCTTACGGCTTCTCTTTGGCGCGCTCGATATTTGGGAACCG 637
DB 532 ATGATGATGGGCTTACAGAGTACAGAGTTCGGGTTCGGCTTCGG--CGCTGACACAGTGGGAGCG 589
QY 638 AAMAGATATCTACTGCTGTCAGCAAAAGAGTGGGTGGGCACTTCTGAGCAACCGCTACG 697
DB 590 ---ATGAGGCTATTGGGCAAGAGCAAGCAACCTGCGT-----CGGCG 628
QY 698 GCGACGCTGAACAGGACAGAGACCATCGAANAACCGCTCGCGGCTGTCCNAATGGCTGTGA 757
DB 729 ATGACGCTTACAGCGTACAGCGATCTGGAAGACCGCGTGGCGCGCTGACAGTGGGGTGA 688
QY 768 TGTATCTGACGCGAAGGGTGTATTAGCGGATGAGCGATTCGCTGAGACCGACACAGCG 817
DB 689 TCTAAGCTCAACGGAGGCGGACGAGCGCAACCGGACCCCAATGGGCGGCGCTCGACNA 748
QY 818 TACTTCAAAACCTTCGCCCTATGATGGGATGACGACCAAAAAACCGCAGCCCTTCACAGCT 877
DB 749 ATCGGAGAGTGTGGGCGCATATGGCCATGGAAGAGCTTGAAACAGAGCGCGCTGATCTCG 808
QY 878 GCGGCGCACACCGCTGGTATTGTGACGATATGGCATGGCTCTGTGGTATGAGCCCTGACC 937
DB 809 GCGGCTCACACTTTCGCTTAGAACCATGCGCGCGCGCGCGCGCTGTGCTGGCGCGCTCG 945

```

```

938 CAAAAGGCTCTGACGTTGAAAAGCAGGGGTAGTGTGGGCAACCCCAACATCGACGGCA 997
DB 869 CGAGGCTGCTCGCTGGAGCAGATGGGCTTGGAGAGCTGTATGGCACCCGAA 928
QY 998 AGCAAGCAACCGCGTACCTCCGGGTATCGAAGTGTCTTGGACCAACCAACCCACGAAT 1057
DB 928 CCGGTAAAGACCGCATACACAGCGGCATCGAGGTGTATGGACCAACACACCGACGAAT 988
QY 1058 TGTGATGGGCTATTTCGAGCTGTCTGGCTAGCATTCGCAATGCGAGTGAAGAAGAGCTG 1117
DB 989 GGGACACAGTTTCTCGAGATCTCTGACGCTACGAGTGGGAGCTGACGAAGAGCCG 1048
QY 1118 CGGTGGCCACCATTTGGGACCGCTTGACATCAAAAAGGAAAAACAGCGGTTGACGCCA 1177
DB 1049 CTGGGCTGTGGCATACACGCGCAAGAGCGCGCGTGGCGGACCAATCCCGAGACCGT 1108
QY 1178 GCGAGCCCTCTATTGCGCACACCGGATCATGCGAGATGCGGATATGCGGATTAAGATA 1237
DB 1109 TCGGCGGCGC---AGGCGCGCTCCCGACGATGCTGGCCACTGACTCTGCTCGCGGTGG 1165
QY 1238 ATCCGACCTATCGGCGCTATCTGCGAAAAATTCATGGCGCATCTGTAGTACTTCAGAAAA 1297
DB 1166 ATGATGATATGAGCGGATACACGCGCTGTGGCTTGGAAACACCCCGAGGATTTGGCGGCG 1225
QY 1298 CTTTGTGAAAGCGTGTGTGACGCTGACGCGACGCTGAGCTGGCGGAAATCAGTTGAT 1357
DB 1226 AGTTTCGCAAGCGCTGTATCAAGCTGATGATCGAGAGATGGTGTCCGTTGCGAGTACC 1285
QY 1358 TGGCGCCGGAAGTCCCGAGAAAGCTGATTTGSCAAGACCCGATTCGCGCAGGTAAACA 1417
DB 1286 TTTGGGCGCGCTGTGTTCCCAAGACAGCCCTGTGTGGCAGATCGGCTCTCGCGGTACGA 1345
QY 1418 C--CGACTACTGCSAAGAGTGTGCAGCAGAAATTCGCAAAAGTGGCTGAGGACT-- 1473
DB 1346 CGACTCTGTGCGCAAGCAGATTGCCAGCTTTAAGAGCCAGATCCGCGCATCGGGATTGA 1405
QY 1474 ----AGTGAGATGCTCCACCGTTGGGACAGTGGCGGCTACTTATGGCGGTTCCGATA 1528
DB 1406 CTGTCTCACAGTAGTGTGACCGGCTTGGGCGCGGCGCTGTGCTGTGCTGTGAGCGCA 1465
QY 1529 TCGCGCGCGGTCTAACGTTGCCCGGATTCGCTTGGCGCCACGACAGAGTGGCAGGCA 1588
DB 1466 AGCGCGCGGCGCCACAGGTTGGTGCATCTGCCCTGCAGCCACAAAGTGGGAGGTCA 1525
QY 1589 AGAGCGGAGCGGCTGGCGAAGAGTGTGAGGCTTACGAGCTGTACGAGTCTGTGCGC----- 1642
DB 1526 AGCAACCGCGAGATCTGCGCAAGGCTATTTCGACCTTGAGAGTTCGAGGATCATCTCA 1585
QY 1643 -----ACACCGCGCTACGATGCGGAGCTGATCTTCTGCGCGGTACGCTAG 1690
DB 1586 CTGCGCGCGGGAACATCAAGAGTGTCTTGGCGGACCTGTGCTGTGCTGTGCTGTGCGC 1645
QY 1691 GATTCGAGAAACCGCGAAGACGAGGTTACGATGTGCGGCTTCCCTTCTCGAAAGGCC 1750
DB 1646 CACTAGAGAAAGAGATCTGCGCGAGCTGGCGCAACAGCTGTGCGCGCTGTACCGCGGCGC 1705
QY 1750 TGTGATGGGAGCGAGCGAGATGAGGCGGCAAGCTCTTTCGACGAGCTGTGCGCGCTGCG 1810
DB 1706 CCGACAGTGGCTGTGCGAGCAACAAACCGGATGAGTATCTTTCGCTGTGAGAGCCCAAG 1765
QY 1811 CGGATGCTTCGCGCACTGGCAGAAAGAGTATGTGGTGAAGCGCGAAGAGATGCTGCG 1870
DB 1766 CAGATGGCTTCGAAAGTACCTCGGAAAGGACACCGTTCGCGCGCGCGAGTACATCGCTGC 1825
QY 1871 TGGATCTGGCAGCGTATGCGCTTACCGCGCCGGAATGACCGTGTCTGGCGGCGGTA 1930
DB 1826 TCGAACAAGGGAACCTGCTTACGCTCAGTGCCTTCGAGTACGAGTGTGCTGTAGTGGCGC 1885
QY 1931 TGGCGCTACTGGGCAACAACTATGTGGCAGCAACACGCGGTATTCACCGATTTGTGAG 1990
DB 1886 TGGCGCTCTCTGGGCGCAAACTACAAAGGCTTTACCGTGGGCGGTGTACCGAGAGGCTCGC 1945

```

QY	1991	CCGAGTGCACCAACGACTTTTTCTTGAACTGACCTGACCGATATGGGAACCACTGGAACG---	204
Dd	1946	AGTCATGTACCACGACTTTCTGCTGAACCTGCTGCACATGGGTATCNACTGGAGCCCT	205
QY	2048	CGGTAGGTAGCACGCCCTAGCAAAATCCGCAACAACCGGTGCGGTGAAGTGSACCG	2107
QY	2006	CCCCAGTAGACAGCGGAGCTACACGGGCAAGATGCGAGTGGCAAGGTGAAGTGSACCG	2065
QY	2108	CCTCGCGGATGGATGATTGCTTGCTTGCACATGGCTACTGCGCTCTCTACGACGAGTGT	2167
Dd	2066	SCAGCGCGCTGGACCTGCTCTCGGGTCCACTCGAAGTTCGCGGGGCTTGTGAGGTCT	2125
QY	2168	ACCCCAGGAGGACATAACCGGAGAAGTTCGTGCAGAGACTTCGTCCCGGCTGGACCAAG	2227
Dd	2126	ATGCGCGCATGATGACCGGACGGCAAGTTCGTGCACNGANTTCGTGCTGCGTGGGACMAGG	2185
QY	2228	TGATCAACGCGCAAGCTTTCGACCT	2252
Dd	2186	TGATGAACCTCGACAGTTCGACGT	2210
 RESULT 15 US-08-228-662-2			
: Application US/08228662			
: Accession No. 588639			
: Patient No. 588639			
: GENERAL INFORMATION:			
: APPLICANT: COCKERILL, FRANKLIN R.			
: APPLICANT: KLINE, JAMES R.			
: APPLICANT: UHL, JAMES R.			
: TITLE OF INVENTION: DETECTION OF ISONIAZID RESISTANT STRAINS			
: NUMBER OF INVENTORS: 3			
: NUMBER OF INVENTOR STATES: 1			
: CORRESPONDENCE ADDRESS:			
: STREET: 3500 IDS CENTER			
: CITY: MINNEAPOLIS			
: STATE: MINNESOTA			
: COUNTRY: USA			
: COMPUTER READABLE FORM:			
: MEDIUM TYPE: Floppy disk			
: COMPUTER: IBM PC compatible			
: OPERATING SYSTEM: PC-DOS/MS-DOS			
: SOFTWARE: PatentIn Release #1.0, Version #1.25			
: CURRENT APPLICATION DATA:			
: PUBLICATION NUMBER: 38/08/228,662			
: FILING DATE: 1994			
: CLASSIFICATION: A15			
: ATTORNEY/AGENT INFORMATION:			
: NAME: WOESSNER, WARREN D.			
: REGISTRATION NUMBER: 30,440			
: REFERENCE/DOCKET NUMBER: 150.123US01			
: TELECOMMUNICATION INFORMATION:			
: TELEPHONE: 612/393331			
: TELEFAX: 612/393306			
: INFORMATION FOR SEQ ID NO: 2:			
: SEQUENCE CHARACTERISTICS:			
: LENGTH: 2221 base pairs			
: TYPE: nucleic acid			
: STRANDEDNESS: single			
: TOPOLOGY: linear			
: MOLECULAR WEIGHT: DNA (genomic)			
US-08-228-662-2			
 Query Match                      25.6%       Score 578.2; DB 1; Length 2221;			
Best Local Similarity    57.8%; Pred. No. 1e-152;			
Matches 1229; Conservative    0; Mismatches 838; Indels    58; Gaps    9;			
QY	160	ACACAGGCTGTGGCGGAGAGGCTTGCACTCGATATTTTGCATACAGACAGTGTG	171
Dd	112	ACACAGGCTGTGGCGGAGAGGCTTGCACTCGATATTTTGCATACAGACAGTGTG	171
QY	220	TGACACCCGATGATCGCGATTTCACACTCGGTGAGAGAGTACGCAAGCTGATTTGAC	279

Db 1226 AGTTCCGACAGGGCTGGTACAAAGTGATCCACCGAGACATGGGTCCGTTGGAGATACC 1285  
 Oy 1358 TCGGCGCGGAGTCCCGCCAGACAGCTGATTTGGCAAGCCGATTCGCGCAGTAA 1417  
 Db 1286 TTGGGCGCGTGTGTCGCCCAAGCAGACCCCTGCTGTGGCAGGATCCGGTCCCTGCGGTGACGA 1345  
 Oy 1418 C--CGACTACTGGGAAGTGGTCAAGCAGAAATTTGCACAAATGGGCTCGAGCAIT-- 1473  
 Db 1346 CAGCTCTGCGCGGAGCAGATTTGCCAGCTTAAGAGCCAGATTCGGGGCATCGGGATGA 1405  
 Oy 1474 -----AGTCAGATGCTCTCCACGCTTGGCAGATGCGCGTACTTATCCGCGTCCGATA 1528  
 Db 1406 CTGTCTCACACTAGTTTGGACGCATTTGGCGGCGGTGCTGCTCTCCGTGTAGCGACA 1465  
 Oy 1529 TGGGCGCGGTGCTAACGGTGGCCCGGATTCGCTTGGCCGCCACAGACAGAGTGGCAGGCA 1588  
 Db 1466 AAGCGCGCGCGCGCCAGCGGTGTGCTGCGCTTCAGCCACAGTCCGCTGGAGGTCA 1525  
 Oy 1589 ACGAGCGGAGCGCTCGGGAAGTGTGAGGCTCTAGGAGCAGATCTCTGCCG----- 1642  
 Db 1526 ACGACCGCGAGGATCTGCGCAGGTCTATTCGCACCTGAAGAGATCCAGAGATCAATTCA 1585  
 Oy 1643 -----ACACCGCGCTAGATCCGCGAGCTGATGCTGTGGCGCGTAGGTAG 1690  
 Db 1586 CTCGGCGCGGACATCAAGTGTCTTCGCGACCTGTGCTGTGCTGTGCTGTGGC 1645  
 Oy 1691 GCATCGAGAAAGCCGCAAGCAGCAGGTAGCATGTGCGCGTTCCTCTCTCGAAGGCC 1750  
 Db 1646 CACTAGAGAAAGCACAAGCGGCTTGGCCACAACATCACGTGGCCCTTCACCCCGGCC 1705  
 Oy 1751 GTGGGATCGGCGCGCGAGATGACCGACAGACTCTCTCGCACCGCTGAGCGCGCTGG 1810  
 Db 1706 CGCAGATCGTCCGACGACCAACCGACGTGATCTCTTTCGCTGTGAGCGCCCAAGG 1765  
 Oy 1811 CGCATGGCTTCCGCACTTGGCGACAGAAGATATGTGTGACGCGSAGAGATGCTGC 1870  
 Db 1766 CAGATGGCTTCGGAATCTACTCTCGAAGAGGCAACGCTTTCGCGCGCGAGTACATCGCTGC 1825  
 Oy 1871 TGGATCTCGCAGCTGATGGGCTTAAACCGGCGCGAATGACCTGCTGCTGGCGGTA 1930  
 Db 1826 TCGAAGCGGACCTTGTATGGCTCATGCTCCCTGAGATGAGGTGCTGGTGGTGGCC 1885  
 Oy 1931 TGGCGGTACTGGCACCACTACTATGTGTGGCACCAACACGCGCTATTCCACCGATTTGAG 1990  
 Db 1886 TCGCGCTCTCGCGCAACTACAGCGCTTACCGCTTGGCGCTGTTCACCGAGGCGCTCG 1945  
 Oy 1991 GCGAGTTACACACGACTTTTGTGACCTGACCGGATATCGGGAACAGCTGGAAC--- 2047  
 Db 1946 AGTCACCTGACACAGCTTTCTGATGCTGCTGACATGCTGACATGAGTCTGCTGGAGCCCT 2005  
 Oy 2048 CGGTAGTATGCAACGCTTACGAATCCGCGACCGCAAGACCGTGGCTGAAGTGGACCG 2107  
 Db 2006 CCGCACGATGACGGGACCTTACCAGGCAAGGATGGCAGTGGCAAGSTGAAGTGGACCG 2065  
 Oy 2108 CTCTCGGGTGGATCTGGTATTTGTTCCAACTCGCTACTCGCTCTTACGCGAAGTGT 2167  
 Db 2066 GAGGCGCCTGGACCTGTCTTGGTCTCAACTTGGAGTTCGGGCGCTTGTGAGGTCT 2125  
 Oy 2168 AGCGCCAGCAGATACCGCGAGAGTGTGTCAGAGACTTCGTGCGCGCTCGACCAAG 2227  
 Db 2126 ATGGCGCGATGAGCGGCGAGGTTCGTGACAGGATTCGTGCTGGGTGGGCAAGG 2185  
 Oy 2228 TGATGAACCGCGACCGTTCGACGT 2252  
 Db 2186 TGTGAACTCGACAGGTTCGACGT 2210

Search completed: October 8, 2003, 15:33:54  
 Job time : 143.677 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: October 7, 2003, 19:31:42 ; Search time 16.4119 Seconds  
(Without alignments)  
2194/727 Million cell updates/sec

Title: US-09-884-889-8  
Perfect score: 4002  
Sequence: 1 MENHSGSYNTTGKCK.....VKDFYKAKAYNDLOREDLK 745

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
5% significance level. The score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	2826	70.6	740	1	Q87864 streptomyces
2	2716	67.9	740	1	CATB_STRE
3	2714.5	67.8	735	1	CATB_BACST
4	2691	67.2	737	1	CATB_CAUCR
5	2564	64.1	749	1	CATB_LEGNP
6	2544	63.6	740	1	CATB_MYCBO
7	2544	63.6	740	1	CATB_MYCBO
8	2441.5	63.5	746	1	CATB_MYCBO
9	2485.5	62.1	753	1	CAT2_NHCR
10	2447	61.1	739	1	CATB_MYCSM
11	2444.5	61.1	726	1	CATB_SALTI
12	2442.5	61.0	726	1	CATB_SALTI
13	2423.5	60.6	726	1	CATB_ECOLI
14	2421.5	60.5	730	1	CATB_HLMA
15	2388	59.7	752	1	CATB_MYCBO
16	2380.5	59.5	737	1	CATB_MYCBO
17	2265	56.6	741	1	CATB_RHOCA
18	2239	56.1	761	1	CATB_RHOCA
19	316	7.9	361	1	CCPR_YEAST
20	316	7.9	361	1	CCPR_YEAST
21	244	6.1	249	1	APX1_ANATH
22	229	3.7	1247	1	APX1_ANATH
23	122.5	3.1	1273	1	UVR4_RHME
24	120.5	3.0	978	1	UVR4_RHME
25	119.5	3.0	315	1	PER4_ANATH
26	119.5	3.0	537	1	GAG_BAFV
27	119.5	3.0	1051	1	TRC4_ECOLI
28	117.5	2.9	1279	1	APU_THESA
29	116.5	2.9	1279	1	APU_THESA
30	116.5	2.9	783	1	YGR_KECOLI
31	113.5	2.8	882	1	OPB1_PSEAE
32	113.5	2.8	882	1	OPB1_PSEAE
33	113	2.8	2551	1	PSP1_BACSU
34	113	2.8	3321	1	PCN2_HUMAN
35	112.5	2.8	1444	1	ADP1_MYCE
36	111	2.8	1173	1	DF3A_PSEFL
37	111	2.8	1173	1	DF3A_PSEFL
38	109.5	2.7	807	1	AFSK_STRGR
39	109	2.7	1246	1	OSTA_ECOLI
40	108	2.7	784	1	NARG_SALTI
41	107	2.7	797	1	AF32_HUMAN
42	106.5	2.7	910	1	STL_NEIMA
43	106.5	2.7	910	1	STL_NEIMA
44	105.5	2.7	1616	1	SLAP_BACST
45	105	2.6	292	1	EFTS_MYCPE

## ALIGNMENTS

RESULT 1  
CATB\_STRE  
ID CATB\_STRE STANDARD: PPT: 740 AA.  
AC Q87864;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Peroxidase/catalase (EC 1.1.1.6) (Catalase-peroxidase).  
GN PDB:  
OS Streptomyces reticuli.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID:1926;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN:145;1840; PubMed:10217489;  
RA Zou P, Borovok I, Ortiz de Lucena D, Muller D, Schrempf H.;  
RT "The mycelium-associated Streptomyces reticuli catalase-peroxidase,  
its gene and regulation by FurS";  
RL Microbiology 145:549-559(1999).  
CC -1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND  
CC -1- BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) + 2 Fe(2+) + 2 H(2)O.  
CC -1- COFACTOR: HINS 2 Prothione IX and 2 iron ions per tetramer (By  
CC similarity).  
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL  
CC PEROXIDASE/CATALASE SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the University of Cambridge. The copyright for the data is held by the  
CC user, by non-profit institutions as long as its content is not  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: Y14317; CAY4698.1;  
CC FSP: P00414; P00414; P00414;  
CC InterPro: IPR000763; Bac\_ctase/prase.  
CC InterPro: IPR000763; Peroxidase.  
CC Pfam: PF00141; peroxidase.1.  
CC PRINTS: P00458; PEROXIDASE.  
CC TIGRfams: TIGR00198; cat\_per-HPI; 1.  
CC PROSITE: PS00435; PEROXIDASE.1; 1.  
CC PROSITE: PS00435; PEROXIDASE.2; 1.  
CC PROSITE: PS00435; PEROXIDASE.3; 1.  
CC Oxidoreductase Peroxidase; Iron; Heme; Hydrogen peroxide.  
CC ACT\_SITE 105 105 BY SIMILARITY.  
CC ACT\_SITE 109 109 BY SIMILARITY.  
CC METAL 272 272 IRON (HEME AXIAL LIGAND).  
CC SEQUENCE 740 AA; 81345 MW; E21860AE4B4A0E CRC64;  
CC

Query Match 70.6%; Score 2826; DB 1; Length 740;  
Match Loc. Similarity 71.4%; Pctid: No. 9.4e-177;  
Matches 531; Conservative 77; Mismatches 126; Gaps 6;





QY 721 SEEFKVDQKAKAKVMDLDFD 744  
 Db 702 NGEFVDFINAWKYANADREDL 725

RESULT 4

CATL\_CACOR STANDARD: PRT: 737 AA.

AD CS1065  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).  
 GN KATG OR CC3043.  
 OS Caulobacter crescentus.  
 OS Caulobacter crescentus.  
 OC Caulobacteriales; Caulobacteriales; Caulobacter.  
 OX NCBI\_TaxID=155892;  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RC MEDLINE=21173698; PubMed=11259647;  
 RA Sherman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Korman W.C., DeLong E.M., Kelley M.R., Ohta M., Maddock J.R.,  
 RA Potocka T., Nelson R.J., Durrin A.S., Slayman D., Ely B.,  
 RA Deboy R.T., Deason R.C., Durkin A.S., Gellm M.L., Harte H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Uterback S.L., Tran K., Wolf A., Vamathevan J., Smolova M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT Complete genome sequence of *Caulobacter crescentus*.;  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 RP SEQUENCE OF 1-494 FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RC MEDLINE=98012985; PubMed=9352936;  
 RA Seelman H.M., Farced F., Weinstein L.;  
 RT Catalase-peroxidase of *Caulobacter crescentus*: function and role in  
 RT stationary-phase survival.;;  
 RT J. Bacteriol. 176:1666-1676(1997).  
 CC -1- FUNCTION: BIFUNCTIONAL EXHIBITING BOTH A CATALASE AND  
 CC -1- BROAD-SPECTRUM PEROXIDASE ACTIVITIES  
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.  
 CC -1- COFACTOR: Binds 2 protholene IX and 2 iron ions per tetramer (By  
 CC similarity).  
 CC -1- INDUCTION: By hydrogen peroxide.  
 CC -1- INDUCTION: BELONGS TO THE PEROXIDASE FAMILY.  
 CC -1- PEROXIDASE (EC 1.11.1.6).  
 CC -1- PEROXIDASE (EC 1.11.1.6).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. Usage by and for commercial  
 CC entities requires a license agreement (see [http://www.isb-sib.ch/announcements/](http://www.isb-sib.ch/announcements)  
 CC or [contact@isb-sib.ch](mailto:contact@isb-sib.ch)).

DR ENBL; A6005967; AAC25005.1;  
 DR ENBL; AF027168; AAC45850.1;  
 DR PIR; A87626; A87626.  
 DR PIR; A45480; A45480.  
 DR RSP; P00431; IRYC.  
 DR RSP; P00431; IRYC.  
 DR InterPro: IPR000763; Bac\_citase/phrase.  
 DR InterPro: IPR002016; Peroxidase.  
 DR Pfam: PF00141; peroxidase; 1.  
 DR PRINTS; PR00458; PEROXIDASE.  
 DR TRIGRAMS; TIGR00198; cat\_per\_HPI; 1.  
 DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
 DR PROSITE; PS00435; PEROXIDASE\_2; 1.  
 DR PROSITE; PS00435; PEROXIDASE\_3; 1.  
 DR Oxidoreductase; Peroxidase, Iron, Heme; Hydrogen peroxide;  
 KW Oxidoreductase; Peroxidase, Iron, Heme; Hydrogen peroxide;  
 KW Complete proteome.  
 ACT\_SITE 102 102 BY SIMILARITY.

FT ACT\_SITE 106 106 BY SIMILARITY.  
 FT METAL 268 268 IRON (HEME AXIAL LIGAND).  
 SQ SEQUENCE 737 AA; 60165 MP; D500742D04E02E97 CRC64;  
 Query Match 67.2%; Score 2691; DB 1; Length 737;  
 Best Local Similarity 69.5%; Pred. No. 5,9e-168; Indels 14; Gaps 6;  
 Matches 508; Conservative 64; Mismatches 145; Indels 14; Gaps 6;

QY 15 NFGKCPGTGSKSAGGCTKRDWMPNNLNLGILQRHSLSDDPNDPFDYAEFFKLD 74  
 Db 18 NLRKCP-----WGIRGPNANRWMPQSLRLGLNHPASNPNGCFATFAEAKSLD 70  
 QY 75 LAAYKDLAALMTDSQWPAADYCHYGPPTMAHSACTGYRCDGCGGCSOSRPA 134  
 Db 71 LDAYVSDJLALMTDSQWPAADYCHYGLFIRLAWHAAGTYRITDGRGAGGCGQREAP 130  
 QY 135 NSPDANLDAKRLMLPDKYGRKISWADIMLTGSVALETGKTFKTFAGGAGDWE 194  
 Db 131 NSPDVNTLDAKRLMLPDKYGRKISWADIMLTGSVALETGKTFKTFAGGAGDWE 190  
 QY 195 PEEDVTGASTELWLDKRYEGDRENLGNLQVOMGLTYVPSGPKGKPTAAANDRET 254  
 Db 191 PEE-LWGPSTWLDKRYEGDRENLGSPGAVOMGLTYVPSGPNGLPLASARDIRET 249  
 QY 255 FORMANDEETVALIAGCTPKTHGAADAERYVGREPAAAGIEMLGKNTYTGCGA 314  
 Db 250 FANWANDEETVALIAGCTPKTHGAAGADG- SLVGVPEGGATGAGQFNARHGTGKP 308  
 QY 315 DTITGSLGAWTKPTTQSNHNFPELNGYEMELTSPAGAYOMKPGKAGAGTIPDAHP 374  
 Db 309 DAITGCPPEVLTQPTTQSNHFPDLNLFKYEMLTQSPAGAKQWAKN--APADIPDAFP 366  
 QY 375 SKSHAFMLTDLALRMDPDYKISRKYENPDEFADAFKAKYKLTHRDGMKPVYGLP 434  
 Db 367 KNTVTFMLDLSALRDEYKISRKYENPDEFADAFKAKYKLTHRDGMKPVYGLP 426  
 QY 435 EYQEDLLWQDDIPDVSHPHLDNDISGLKAKILBSGLTYVSLVYASNASASTPNSDKR 494  
 Db 427 LVPKEELIWQDDIPAVDHPHLDNDISGLKAKILATGLSADLVSTANASASTYQSDKR 486  
 QY 495 GGANGARILAQKQMEVYNPQGLARVLTLEGLQEDPQASDNKAVSLADLVLAGCA 554  
 Db 487 GGANGARILAQKQMEVYNPQGLARVLTLEGLQEDPQASDNKAVSLADLVLAGCA 546  
 QY 555 GYERAAKDAKHGVQFVFNGRDADATSDTDEAFALPAADGPRNVLKP-ZHKVSABEM 613  
 Db 547 ATEKAAKDAKTSVFPFAPGMDASADTDAHFEALPESDPSGPRNGRHYMAPEA 606  
 QY 614 LVPKALTSKPTVATLQVGMVLTGNTDQSGVFTNKPGQISNDFFVNLIDNTKWR 673  
 Db 607 LVDRAGLLSGPELVATLQVGMVLTGNTDQSGVFTNKPGQISNDFFVNLIDNTKWR 665  
 QY 674 ASDSKYFGKDFKFGVKNSTGIVDLTIGNSNLSRALAEYVCGASERKFKVDFVYKAW 733  
 Db 666 -SPTAANAFAGHDKRSSEPRMTATVOLLFGSHALRAFAEYVACASQERKFKVDFVYKAW 724  
 QY 734 AKVQDLSREDL 744  
 Db 725 NKVWANDRLDL 735

RESULT 5

CATL\_DGPN STANDARD: PRT: 749 AA.  
 AC CSW456  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).  
 GN KATG.  
 OS Legionella pneumophila  
 OS Legionella pneumophila  
 OC Legionellaceae; Legionella.



```

Query Match          63.6%; Score 2544; DB 1; Length 740;
Rec'd Local Similarity      84.3%; Pred. No. 2.e-158;
Matches 480; Conservative 88; Mismatches 160; Indels 10; Gaps 8;

QY 2 ENKHSSSTNTNTNGCPTTGGSLAQAGGCTKRNMPNMLNLGLTLQHSSLSDPND 61
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 3 EOHPTTETTTGAASNG-CPVVGKHKYVPEGGS--NODWPNRLNLKLVHQNPAVDPMG 59
QY 62 PDYDYAEFFKLLDLAANKYDLDLALMTSDQNPADYGHGYPFFIRMAHSAAGTYRDCOR 121
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 60 AEFDYAEAVATIDVALTRDLEWTTSQWPNADYGHGYPLEINMAMAGTYRIDROR 119
QY 122 GGGSCSORPAPLSPDNANLMDKARLLLPITQKYGKISWADMLITGNVALFTGFK 181
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 120 GAGGQGHQRPAPLSPDNANLMDKARLLLPVKKYKGLSWADLVFANGCALSSGFK 179
QY 182 FGAGGQGHQRPAPLSPDNANLMDKARLLLPVKKYKGLSWADLVFANGCALSSGFK 241
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 180 TFGGFGVDPNDEPDE-VYNGKEATVGLYSGKRDLENLPAVQKGLIYNVPAENGN 238
QY 242 PDPTAAARDIRTEGRMANDEETVALTAGGRTFGKTHGAAQAKYVGRPEAAAGIEMS 301
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 239 PDPMANAAVDIRTEGRMANDEETVALTAGGRTFGKTHGAAQAD-LVGPEPEAPLQMG 297
QY 302 LQKNTYCTGCGANTTSGLEPGANTYCTPONSSENFENLEKLETSRAGNKKRPTD 361
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 298 LQKNSYCTGCTGKADITSGIVETWNTPTKNDNSFLELLYGVENELTSPAGAKYTKD 357
QY 362 GAGAGTIPDAH-DPSKSHAPFNLCTDIALRMDPOYEKTSRRYENDEDFADAFKMYKL 420
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 358 GAGAGTIPDPGGGPRSS--PMLNATDLSURVDYPTERTLGEHEPEELADEFAKMYKL 415
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 412 FREDMGVYKGLPEYDPOEDITWDDPTDPSVHPLVDENDLEGLKAKLLESSLTVSELP 480
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 416 THRDMPVARYLGLPVKQTLNQDPPVAVSHDLVGEAEIASLQSLASSLTVSLSVT 475
QY 481 AMASATSRNSKGGANGARIELAFOKQEVNFPQ-OLARVLTLEGIEDFENQAOQSDN 539
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 476 AMAAASPSGSDKGGNGGRIKQPOGVENVDPDGLKRTLEELIESPSNAAQGN 535
QY 540 KAVSLADILVLACGAGVEKADAGHGVOPVPPNGRADATACQDVEAFALPEADGPR 599
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 536 IKVSPFADILVVGCAALEKAAKAAGHNTVPTPGRTDSQSDQDVESEFVLEPKADGPR 595
QY 600 NYIKPKHVGAEMLDVRAQLLSAPENTALVGNRVLTGYDSQGVFTNKKPOLSN 659
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 596 MTLKGNPLPFLATLWAKSDNKLPEKRTGVYGLVGLVGNATRLPGLVFAESLVTG 655
QY 660 DFFVNLNDLWAKSDNKLPEKRTGVYGLVGLVGNATRLPGLVFAESLVTG 719
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 656 DFFVNLNDLWITWEPSPADQDTQSGK-VGSKVKMTGSRVGLVFGNSRLALIVEYGD 714
QY 720 DSEKVPVDFVMAKMYKLDORPLK 745
Dd : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dd 715 DAQPRFQDPAVNDKYNKLDREVR 740

```

## RESULT 7

CATA\_MCTCU

ID CATA\_MCTCU STANDARD; PRT: 740 AA.

NC 00812; 00821; 050544; 050546; 050551; 050552; 050553; 050554;

AT 01-OCT-1994 (Rel. 30.4; Created)

DT 30-MAY-2000 (Rel. 39; Last sequence update)

DT 28-FEB-2003 (Rel. 41; Last annotation update)

DE Peroxidase/catalase T (EC 1.11.1.6) (Catalase-peroxidase T).

GN KATG OR RV1908C OR MT1959 OR MTCY180.10.

OS Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;

OC Mycobacteriaceae; Mycobacteriaceae; Mycobacterium.

OX NC\_018412-173;

RN (1)

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC STRAIN-H37Rv; MEDLINE=9308108; PubMed=8320241;  
 RX Heym B., Zhang Y., Soule S., Young D., Cole S.T.;  
 RA "The catalase-peroxidase gene and isoniazid resistance  
 RT required for the isoniazid susceptibility of Mycobacterium  
 tuberculosis."; J. Bacteriol. 175:4255-4259(1993).  
 RL [2]  
 RN REVISIONS.  
 RP STRAIN-H37Rv;  
 RC SOLC S.T.; Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 25618;  
 RA Cockrell F.R., Ohl J.R., Temesgen Z., Zhang Y., Stockman L.,  
 RA Roberts G.D., Williams D.L., Kline B.C.;  
 RT Rapid identification of a point mutation of the Mycobacterium  
 tuberculosis "catalase-peroxidase" (katG) gene associated with  
 isoniazid resistance"; J. Bacteriol. 175:4255-4259(1993).  
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-INH-resistant strains;  
 RA Martilla H.-J., Soini H., Huovinen P., Villanen M.K.;  
 RT Identification of isoniazid-resistant Mycobacterium  
 tuberculosis strains isolated from Finland; J. Clin. Microbiol.  
 32:103-106(1994).  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-H37Rv;  
 RC MEDLINE=98295987; PubMed=9634230;  
 RX Cole S., Brech R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Berrington J., Brown D., Brown C.E., Chillingworth T., Cohen P.,  
 RA Badcock K., Basham D., Bason M., Bingle C., Bingle P.,  
 RA Davies R., Deakin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,  
 RA Horsley T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sutton J.B., Taylor K., Whitehead S., Barrett B.G.;  
 RT The genome sequence of Mycobacterium tuberculosis H37Rv  
 RL complete genome sequence"; Nature 393:537-544(1998).  
 [6]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-CDC 1551 / Oshkosh;  
 RA Fleischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson R., Dodson R., Gwinn K.L., Hart D., Hickey E.,  
 RA Karp P., Ketchum K., Kohn M., Kozminski D., Kozminski S.L.,  
 RA DeLcher A., Uitterback T., Feldman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains"; Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN PARTIAL SEQUENCE FROM N.A., AND ISONIAZID RESISTANCE.  
 RP STRAIN-H37Rv;  
 RC MEDLINE=92365822; PubMed=1501713;  
 RX Zhang Y., Heym B., Allen B., Young D., Cole S.T.;  
 RA "The catalase-peroxidase gene and isoniazid resistance of  
 RT Mycobacterium tuberculosis"; Nature 358:591-593(1992).  
 RL [2]  
 RN SEQUENCE OF 1-94 FROM N.A.  
 RP STRAIN-H37Rv;  
 RA Song J., Deretic V.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND  
 CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES. MAY PLAY A ROLE IN THE  
 CC INTRACELLULAR SURVIVAL OF MYCOBACTERIA.  
 CC -1- CATALYTIC ACTIVITY: 2 H2O(2) - O(2) + 2 H(2)O.  
 CC -1- COFACTOR: 2 Fe atoms, 2 protons, 1x and 2 iron ions per tetramer.  
 CC -1- SUBUNIT: Homodimer (Probable).









```

Db 478 VSTAKASTFSSDGGANGRIQLQGLWEANEPEDELAQWVKYEIQ-----NAGS 533
QY 538 DNKAVSLADLVLAGCAGYKAADGHEVYVPPENGRADATQEDTFALEPAAGD 597
Db 534 IN--VSFADVLVSGNGYVYKAAGAAGFDVVPTEFGDGTQEDTVDVDFSLPFPADG 591
QY 598 FNTYKPEHKVSAEMLVYDRAQLLSAPEMTALVGNHVLGTVDGSGHGVTKRQGL 657
Db 592 SNGFVNLDDNPEFLKIDRANLGLGAPMTILVGLVLDVWHGQTKRGVLTDPGAL 651
QY 658 FNTFVNLDDNWKRAEDSVFEGSDPKTEGVNSGTRVDLFGNSLSALAEVYG 717
Db 652 TDFVNLDDNKTAKFPADDTYIGTDRATGSPKATGTRVDLVFASNSQLAEVYA 711
QY 718 CADSEKTVKVFVYKANKYMDLDFDL 744
Db 712 EDDSKREYVDFVYKANKYMDNDYDV 738

RESULT 11
ID CATL_SALTI STANDARD: PRT: 726 AA.
AC Q82303: 2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Peroxidase/Catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)
DE (Hydroperoxidase I).
GN KATG OR STY3760 OR T3510.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID:601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RC Kirschill C., Dougan G., James K.D., Thomson N.R., Pichard D., Main J.,
RC Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RC Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RC Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RC Krogh A., Larsen L.S., Leathers S., Moulé S., O'Garra P., Parry C.,
RC Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RC Wainwright S., Wardle J., White O., Wilson A.C., Young I.M.,
RC Zdobych E., Barrett B.G. & a multiple drug resistant Salmonella
RC enterica serovar Typhi CT18.
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TY2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RC Deng M., Liou S.-R., Plunkett G., III, Mayhew G.F., Rose D.J.,
RC Kaper J.B., Tarr P.I., Tarr P.I., Tarr P.I., Tarr P.I., Tarr P.I.,
RC Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RC and CT18.
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- FUNCTION: Bifunctional, exhibiting both a catalase and
CC -1- broad-spectrum peroxidase activities (By similarity).
CC -1- CATALYTIC ACTIVITY: 2 H2O2(2) = O2(2) + 2 H2O.
CC -1- COFACTOR: Sulfide 2 protoheme IX and 2 iron ions per tetramer (By
CC -1- SUBUNIT: Homotetramer (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC -1- SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at EMBL Grenoble. The sequence data have been deposited in the EMBL
CC database. The sequence data have been deposited in the EMBL database
CC use by non-profit institutions. The sequence data have been deposited in the
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sb.ch).
DR EMBL: AL627279; C400915.1.1.
DR EMBL: AE016846; A071018.1.1.
DR InterPro: IPR000763; Bac_c1ase/prase.
DR Pfam: PF00141; peroxidase.
DR PRINTS: P00458; PEROXIDASE.
DR TIGRPFAM: TIGR00198; cat_per_hpi.1.
DR PROSITE: PS00436; PEROXIDASE.1.
DR PROSITE: PS00436; PEROXIDASE.1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 102 102 BY SIMILARITY.
FT ACT_SITE 106 106 BY SIMILARITY.
FT METAL 267 267 IRON (HEME AXIAL LIGAND).
FT REPEAT 15 20 HRM.
SQ SEQUENCE 726 AA; 8701C95343EEB38 CRC64;
Query Match 61.18; Score 2444.5; DB 1; Length 726;
Best Local Similarity 63.68; Pred. No. 76:152;
Matches 468; Conservative 83; Mismatches 166; Indels 19; Gaps 7;
QY 11 TYNTNGKCPPTGSGSKDSAGGCTKNRMWPNMLZILUHQHSLSDNDPDDFAEF 70
Db 7 THNTLSTGCPFHGGHDSAGACTASQNPQLNOLLNQHNSNPILGDFDIFGE 66
QY 71 KLLDLAAVKDLAALMTDSQDMPADYGHYGFPIHMAHWSAGTYRIGDGRGGSGSOR 130
Db 67 SKLDYSALKDGLKALLTDSQDMPADNGSYGLFIHMAHIGAGTYRISIDRGAGRQOR 126
QY 131 FAPLNSFDNANDKARLLMPTIKQYKTKISHADLMLTGWVALETGKFTGFPAGRA 190
Db 127 FAPLNSFDNANDKARLLMPTIKQYKTKISHADLMLTGWVALETGKFTGFPAGRA 186
QY 191 DWPEPEEDVYGAETELGDKRYEGDRELENFYGVMGLYVNPSPGSDPTAAARD 250
Db 187 DWPEPEEDVYGAETELGDKRYEGDRELENFYGVMGLYVNPSPGSDPTAAARD 244
QY 251 TRFTFGWMMNDEETVALIAGCTTGCTGGGAADENKYVGFEDPAAGIENSLGHVNYGT 310
Db 245 TRATFGWMMNDEETVALIAGCTTGCTGGGAADENKYVGFEDPAAGIENSLGHVNYGT 303
QY 311 GHGADTYTSLGEGAMTKTPTQNSNFNFLENFYENELTKSPAGAYQWKPQAGAGTIPD 370
Db 304 GVGADATYSLGELVWYTPPTQNSNFNFLENFYENELTKSPAGAYQWKPQAGAGTIPD 361
QY 371 AHDEPESHANPPTDIALNDPQYKTSRYENVPEDFAKAKATKLTIRHDGQKVR 430
Db 374 AHDEPESHANPPTDIALNDPQYKTSRYENVPEDFAKAKATKLTIRHDGQKVR 421
QY 362 PPDFSKKPKTMLVTLDTLRFDEPEKISRRELNDFDQAFNEAFARANKFLTHRDGPKAR 421
QY 431 YLGSFVPPDRLTWQDPTDPVSHPLVD--ENDEGLKAKLTESGLTVELSVTAWSASTF 488
Db 422 YLGSFVPPDRLTWQDPTDPVSHPLVD--ENDEGLKAKLTESGLTVELSVTAWSASTF 478
QY 489 RNSDKGGAGARLARLPADKQWYVNPQOLARVLKTLGTDGDSNQAQSNKAVSLADLI 548
Db 479 RGSDDKGGAGARLARLARPORWEVN--AVALRVLPVLEALQKTENKA-----SLADI 529
QY 549 VLAGCAGYKAADGHEVYVPPENGRADATQEDTFALEPAAGDPTFPIPKHVK 608
Db 530 VLAGCAGYKAADGHEVYVPPENGRADATQEDTFALEPAAGDPTFPIPKHVK 599
QY 609 SAPEMLVDRAQLLSAPEMTALVGNHVLGTVDGSGHGVTKRQGLSDNFDVLDL 668
Db 599 TTESLLIDKAQQLTLAPMTVLGVRVLGTNFDGSGVQVTDRRGVSLTDFDFAILLDM 649
QY 669 NTKWRAEDSVFEGSDPKTEGVNSGTRVDLFGNSLSALAEVYGCADSEKTVKD 728
Db 650 NTKWRAEDSVFEGSDPKTEGVNSGTRVDLFGNSLSALAEVYGCADSEKTVKD 709
QY 729 FYKANKYMDLDFDL 744

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sb.ch).



ID CATA\_ECOLI STANDARD; PRT; 726 AA.  
 IC 01-JAN-1990 (Rel. 13, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)  
 DE (Hydroperoxidase 1).  
 GN KATG OR B3942.  
 OS Escherichia coli.  
 OS taxonomic classification: Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OC NCBI\_TaxID=562;  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RA MEDLINE=88314956; PubMed=3045098;  
 RA Triggs-Raine S.L., Bobie B.W., Mulvey M.R., Sorby P.A., Loewen P.C.;  
 RT "Analysis of the Escherichia coli genome. III. DNA sequence of the  
 RT nucleotide sequence of katG, encoding catalase HPI of Escherichia  
 RT coli." Nucleic Acids Res. 21:3391-3398(1993).  
 RN [2]  
 RN SEQUENCE OF 1-339 FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RC MEDLINE=93347969; PubMed=8346018;  
 RA Plunkett G., III, Burland V.D., Daniels D.L., Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. III. DNA sequence of the  
 RT nucleotide sequence of katG, encoding catalase HPI of Escherichia  
 RT coli." Nucleic Acids Res. 21:3391-3398(1993).  
 RN [3]  
 RN SEQUENCE OF 309-726 FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RC MEDLINE=94089392; PubMed=8265357;  
 RA Blattner F.R., Burland V.D., Plunkett G., III, Sofia H.J.,  
 RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the  
 RT region from 89,2 to 92,8 minutes." Nucleic  
 RT Nucleic Acids Res. 21:5408-5417(1993).  
 CC -1- FUNCTION: Bifunctional, exhibiting both a catalase and  
 CC broad-spectrum peroxidase activities.  
 CC -1- CATALYTIC ACTIVITY: 2 H<sub>2</sub>O<sub>2</sub> = O<sub>2</sub> + 2 H<sub>2</sub>O.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- INDUCTION: By hydrogen peroxide.  
 CC -1- PTM: The N-terminus is blocked.  
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL  
 CC PEROXIDASE/CATALASE SUBFAMILY.  
 CC -1- SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.  
 CC This SWISS-PROT entry is copyright. It is produced through collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: W21516; AA234040.1;  
 CC EMBL: L13201; AA303074.1;  
 CC EMBL: U00005; AAC43048.1;  
 CC EMBL: AE000468; AAC76924.1;  
 CC PIR: A65201; CSECHP.  
 CC HSP: P00431; ICYF.  
 CC SWISS-2DPAGE: P13029; COLI.  
 CC Triggs-Raine S.L., Bobie B.W., Mulvey M.R., Sorby P.A., Loewen P.C.;  
 CC Triggs-Raine S.L., Bobie B.W., Mulvey M.R., Sorby P.A., Loewen P.C.;  
 CC InterPro: IPR000763; Peroxidase.  
 CC Pfam: PF00141; Peroxidase; 1.  
 CC PRINTS: PF00458; PEROXIDASE.  
 CC TIGRfam: TIGR00198; cat\_per\_HPI; 1.  
 CC PROSITE: PS00435; PEROXIDASE.1; 1.  
 CC PROSITE: PS00436; PEROXIDASE.2; 1.  
 CC PROSITE: PS00437; PEROXIDASE.3; 1.  
 CC PROSITE: PS00438; PEROXIDASE.4; 1.  
 CC OXfam: OXfam00000; Peroxidase; iron; Heme; Hydrogen peroxide;  
 CC Complete proteome.

FT ACT\_SITE 102 102 BY SIMILARITY.  
 FT ACT\_SITE 106 106 BY SIMILARITY.  
 FT METAL 257 257 IRON (HEME AXIAL LIGAND).  
 FT CONFLICT 631 631 A -> G (IN REF. 1).  
 SQ SEQUENCE 726 AA; 80023 MW; 24D32EBD5E9B6 CRC64;  
 Query Match 60.6%; Score 2423.5; DB 1; Length 726;  
 Best Local Similarity 63.6%; Pred. No. 1.6e-150;  
 Matches 467; Conservative 78; Mismatches 172; Indels 17; Gaps 7;  
 QY 12 YNNTGCKPCTGSGSLKQSGGCTKAKRWKMLGILRHSLSLSDNDQDYAFSEFK 71  
 DB 8 HNTATCKPCHQGHGDSAGAGTTTRDWNQNLAVDLLNHSRNLCEGDFDYKEFS 67  
 QY 72 KLDLAAYKDLAALMTSDQMPADTGHYGFPIFMAHMSAGTIRGDGSGGSGSQR 131  
 DB 68 KLDYGLKDKLALITQSGMPADMGSYAGLFPIFMAHMGATYHSIDRGGAQGR 127  
 QY 132 APLNSWPNANLKDARLLMPTKQYKPTISWADLMTLGVVALETFKGTGEGAG 191  
 DB 128 APLNSWPNVSLDKARLLMPTKQYKPTISWADLMTLGVVALETFKGTGEGAG 187  
 QY 192 VWPEEDYTWGAETEMLDKARYEDRELEPLGAVQMGILTYNPGPNKPPAAARD 251  
 DB 188 VWPEEDYTWGAETEMLDKARYEDRELEPLGAVQMGILTYNPGPNKPPAAARD 245  
 QY 252 RETFCGMANDDETVALLAGHTFGTKGADAKRYVGBEPAAGTPEMSLQWNTY 311  
 DB 246 RATFCGMANDDETVALLAGHTFGTKGADAKRYVGBEPAAGTPEMSLQWNTY 304  
 QY 312 HGADYITSGLEGAMTKTPTGNSNFFENFGYEWELTKSPAGAYOKPKMDGAGAT 371  
 DB 305 VGADATISGLEVMQVTPQNSNFFENFGYEWELTKSPAGAYOKPKMDGAGAT 362  
 QY 372 HDPSKSHAPMLTDLALRMDPDYKISRVENPDEFADAKANFKLTREDMGPKVY 431  
 DB 363 FDPKSKRKPMLVTLTRFDEPEKISRRELPNDPOAFNEAFANFKLTREDMGPK 422  
 QY 432 LGPEVPODLTQDDIPD-VSHPLVDNDLEGLKAKLTESGLTVSELVSTAMASST 490  
 DB 430 LGPEVPODLTQDDIPD-VSHPLVDNDLEGLKAKLTESGLTVSELVSTAMASST 480  
 QY 491 SDKRGKARIRLAPQDMWNNPQOLRLVLTGICIGDFNQASDNKAVSLADLV 550  
 DB 481 GDXRGKARIRLAPQDMWNNPQOLRLVLTGICIGDFNQASDNKAVSLADLV 531  
 QY 551 AGCAVKKAKAGHGVQFPPGRADATAEQTVFAEALPAADGFRNTKPEKHYSA 610  
 DB 532 AGVGVKAKAGHGVQFPPGRADATAEQTVFAEALPAADGFRNTKPEKHYSA 591  
 QY 611 EEMLVDAQLLSISAPENTALLGVGMVLCNTYDGSQGVFTNPKQSLNDPVMJL 670  
 DB 592 EEMLVDAQLLSISAPENTALLGVGMVLCNTYDGSQGVFTNPKQSLNDPVMJL 651  
 QY 671 KWRASDSKVFGRDPTGCVKNGSKTGVOLIFGSNSLELRLAEVYGCADSEKPK 730  
 DB 652 KWRASDSKVFGRDPTGCVKNGSKTGVOLIFGSNSLELRLAEVYGCADSEKPK 711  
 QY 731 KANAVVMDLDRFDL 744  
 DB 712 KANAVVMDLDRFDL 725  
 RESULT 14  
 CATA\_HALMA  
 ID CATA\_HALMA STANDARD; PRT; 730 AA.  
 AC 059651;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Peroxidase/catalase (EC 1.11.1.6) (catalase-peroxidase).  
 GN PERA.



```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@sib-sib.ch).
-----
DR ENBL; AC005159; AAC20931.1;
DR ENBL; AF065761; AAC23534.1;
DR PIR; T44562; T44562.
DR HSP; P00431; IRYC.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR TrEMBL; P00431; Peroxidase.
DR PRINTS; PR00416; Peroxidase.
DR TIGRfam; TIGR00198; cat_per_HPT; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PROSITE; PS00873; PEROXIDASE_4; 1.
DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide; Plasmid;
KW Complete proteome. 70
KW AC-SITE 83 BY SIMILARITY
KW ACT-SITE 83 BY SIMILARITY
FT METAL 248 IRON (HEME AXIAL LIGAND).
SQ SEQUENCE 720 AA; 80476 MW; F88823BCD3CB2F9 CRC64;
-----
Query Match 60.0%; Score 2401.5; DB 1; Length 720;
Resc Local Similarity 59.7%; Pred. No. 4.4e-149;
Matches 447; Conservative 104; Mismatches 165; Indels 33; Gaps 4;
QY 1 MENHKIS-GSSTYNTNGKCPPTGGSLKQSGGCTKRDWPKNLGLTQRHSLSDP 59
DB 1 MENEHNGTS-----DWNPNOLDLELDQNSQOQDP 32
QY 60 NDPPDYAEFKKLDJAAVKOLALMTSDQMPADYGHYQFFIRMAHNSAGTYRIGD 119
DB 33 YEDPDYAEAFEDJLJAAVKODLEKMTSDQMPADYGHYGLPIFRMAHNSAGTYRIGD 92
QY 120 GRGGGSGSOREAPLNSPDNLDKARLLWPTKQYGRKTSWADLMTLGNVALETWG 179
DB 93 GRGGAAGGQRLPPVDSMPDNVNLKARLLWPTKQYGRKLSHGLDILLAGNVALESNG 152
QY 180 FTFGFGAGRADWPEDEYVYGAFTEN---LGDNRYEGRLENPLGAVQMLTYNPE 236
DB 153 FEYGFGRGKDYTFDENYRDPEDDETTSGD-RFDAGSLAWPLAGNYGLTYNPE 211
QY 237 GPNGKPDFIAARIDRTETFGMAWDEBTVALAGHTPKTGHGAADAERYGPEPAAG 296
DB 212 GPNGEPDLEGSARKNIRESEFGKAMMDKETVALAGSHFTGKVGADDPENYGAEPAAAP 271
QY 297 IDEMSLGNKMYTGGHGAADTTISGLEGAMTKTPTQSNNEFFENFGYEMELTSPAGAYQ 356
DB 272 IFAQGLQENHNEFGSGKGFDTTISGLEGPNTTTPQMDSTYDNLLETMEPEKPGGAWQ 331
QY 357 WKPKDCACTIPDAPDPSKSHAPMLTJTLALRMDPDYKTSRYRYENDEFAFAKA 416
DB 332 WTKSGELNESAPQVQDPTEDVNMHTDVALKDDPDYREVLETFQENPREFQCSFKA 391
QY 417 KTKLRDQPKVRYLGEVVPQDLIVQDP:IPVSHPLVDENDIEGLKAKILESGLTVSE 476
DB 392 WTKLRDQPKVRYLGEVVPQDLIVQDP:IPVSHPLVDENDIEGLKAKILESGLTVSE 451
QY 477 LVSTAWASASTFRNSDKGGANGARIRLAPQKQWVNNPOOLARVLKTLGIDQEDNQAO 536
DB 452 LVKTAWASATYRDSKGGANGARIRLEPQSWNEVNEPEQIEAALSTYEDIOAEFNDAK 511
QY 537 SGNKAVSLDLVLACGAVKAKDAGHEVVPFNGRADTAQTOVFAEALPAAAD 596
DB 512 SDGNKAVSLDLVLGNAALTEQAAADAGTYDVFPEFQRTDTPQIDTVESFEALAFRAD 571
QY 597 GF8NYIKPEHKYAEEMLVDRQALLSISAPMTALVGNRVLGNTNYXGSGRGVFTNPKQ 656

```

```

-----
DB 572 GF8NTLGDARREPELLVDRAELLNUTADMTVLVGLRAGLVTHQDESLGIFTDQPT 631
QY 657 LSNDFVWLLDNTWTHASDSKQDFGRDFTGCTGKSGTGYDVLITGNSSELALARYV 716
DB 632 LTNDFFTLLDMQYEMEAASEDEVEFELRDRETGCVETGSRVDLLFGSNTLRALARYV 691
QY 717 GCADSEEEKFYKDFVKAMAKVMDLDRFLK 745
DB 652 GSDADEELFVQDFVDTWSEVMKLRDFLE 720
-----

```

Search completed: October 7, 2003, 19:43:04  
Job time : 20.4119 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:31:42 : Search time 16.5881 seconds  
(without alignments)  
2134.727 Million cell updates/sec

Title: US-09-884-889-6

Perfect score: 4058

Sequence: 1 MNASADLLHSLQQRCRAF.....RDFVAATKYNNADRFVAS 753

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot\_41:\*

Pred. No. is the number of results predicted by chance to have a  
P-value of the test statistic as low as or lower than the observed  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2489	61.3	735	P14412	bacillus
2	2216.5	54.6	739	CATA_MYCSM	Q59557 mycobacteri
3	2168.5	53.4	740	CATB_STREX	O87864 streptomyc
4	2162.5	53.3	737	CATA_CAUCR	O31066 cathecolact
5	2149.5	53.0	740	CATB_STPRO	Q97159 streptomyc
6	2095	51.6	726	CATA_ECOLI	P13029 escherichia
7	2086	51.3	726	CATA_SAUPY	P17750 salmonella
8	2086	51.3	726	CATA_SAUPY	O82553 salmonella
9	2063.5	50.9	746	CATL_MYCIT	Q92553 salmonella
10	2050	50.5	752	CATL_MYCFO	O08404 mycobacteri
11	2042.5	50.3	740	CATA_MYCFO	O08129 mycobacteri
12	2036.5	50.2	740	CATA_MYCBO	P46817 mycobacteri
13	2003	49.4	753	CAT2_NEUCH	Q9x182 neobactera
14	2002	49.3	749	CATA_LEGNN	Q9xw39 legionella
15	1999	49.3	749	CATA_LEGNN	O59651 haemophilus
16	1930	47.6	720	CATA_HLMA	O92553 salmonella
17	1922.5	47.4	737	CATA_VERYE	Q9x182 neobactera
18	1901.5	46.9	741	CATA_ARFPU	O28050 archaeoglob
19	1642.5	40.5	576	CATA_RHOCA	P37743 rhodobacter
20	267	6.6	249	APX1_ARATH	Q05431 arabidopsis
21	248.5	6.1	361	CCPR_YEAST	P00431 saccharomyc
22	245.5	6.0	249	APX1_PEA	P48534 plasm activ
23	245.5	6.0	249	APX1_PEA	Q87750 vibrio para
24	245.5	6.0	249	APX1_PEA	Q87750 vibrio para
25	117	2.9	358	P12_ARATH	O92553 salmonella
26	117.5	2.8	372	TL29_LXCES	Q97406 streptococ
27	113.5	2.8	372	LIGC_TRAVE	P20013 trameetes ve
28	113.5	2.8	1424	NC03_HUMAN	Q9y699 h nuclear r
29	111.5	2.7	485	PE28_ARATH	O9sae67 arabidopsis
30	111.5	2.7	844	VG14_BPMD2	O64207 mycobacteri
31	109	2.7	844	PAC_KLUCCI	P07941 kluyvera ci
32	108	2.7	844	PAC_KLUCCI	Q8tk6 methanosarc
33	108	2.7	844	PAC_KLUCCI	P43052 bacillus li
34	106	2.6	806	1	Q86877 arabidopsis
35	106	2.6	879	1	Q9Ks58 bacillus ha
36	105	2.6	879	1	Q02934 clostridium
37	105	2.6	879	1	Q02934 clostridium
38	105	2.6	879	1	Q02934 clostridium
39	105	2.6	879	1	Q02934 clostridium
40	104.5	2.6	879	1	Q02934 clostridium
41	104.5	2.6	879	1	Q02934 clostridium
42	104	2.6	879	1	Q02934 clostridium
43	103.5	2.6	879	1	Q02934 clostridium
44	103.5	2.6	879	1	Q02934 clostridium
45	101.5	2.5	348	1	Q02934 clostridium

## ALIGNMENTS

RESULT 1	CATA_BACST	STANDARD;	PRT;	735 AA.
ID	P14412;			
AC	01-JAN-1990 (Rel. 13, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DE	28-FEB-2003 (Rel. 41, Last annotation update)			
OS	PEROXIDASE/CATALASE (EC 1.11.1.6) (Catalase-peroxidase).			
CS	PEROXIDASE/CATALASE			
CC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.			
OX	NCBI_TaxID=1422;			
RM	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RP	STRAIN-ATCC 8005 / IAM11001;			
RA	MEDLINE-8959121; PubMed-2670897;			
RL	Cloning nucleotide sequence and expression in Escherichia coli of			
RL	the Bacillus stearothermophilus peroxidase gene (perA);			
RL	J. Bacteriol. 171:4871-4875(1989).			
RM	[2]			
RP	REVISIONS.			
RA	Trakulniamseai S., Alhara S., Miyai K., Suga Y., Yomo T., Negoro S.,			
RL	Trabe I.; (FEB-1992) to the EMBL/GenBank/DBSJ databases.			
RL	Submitted (FEB-1992) to the EMBL/GenBank/DBSJ databases.			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-99116998; PubMed-9920270;			
RA	Matsura T., Miyai K., Trakulniamseai S., Yomo T., Shima Y.,			
RA	Miki S., Yamamoto K., Urabe I.;			
RT	Evolutionary molecular engineering by random elongation			
RL	Nat. Biotechnol. 17:58-61(1998).			
CC	FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND			
CC	BROAD-SPECTRUM PEROXIDASE ACTIVITIES.			
CC	-1- CATALYTIC ACTIVITY: 2 H(2)O(2) - O(2) + 2 H(2)O.			
CC	-1- COFACTOR: Binds 2 protome IX and 2 iron ions per tetramer.			
CC	-1- SUBUNIT: Homodimer.			
CC	-1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL			
CC	PEROXIDASE SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	-----			
CC	EMBL: M28976; AAA22655.1; --			
DR	EMBL: AB020234; BAA37114.1; --			
DR	PIR: JG0520; JG0520.			
DR	HSSP: P00431; 1bJ9.			
DR	InterPro: IPR000763; bac_ctase/prase.			
DR	InterPro: IPR002016; Peroxidase.			
DR	Fam: PF00141; peroxidase; 1.			

DR PRINTS: PRO0458; PROXIDASE.  
 DR TIGRUS: 19021930; HPT: 1.  
 DR PROSITE: P500435; PROXIDASE 2; 1.  
 DR PROSITE: P500436; PROXIDASE 2; 1.  
 DR PROSITE: P50873; PROXIDASE 4; 1.  
 DR PROSITE: P50873; PROXIDASE 4; 1.  
 DR KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.  
 FT ACT\_SITE 97 97 BY SIMILARITY.  
 FT ACT\_SITE 101 101 BY SIMILARITY.  
 FT ACT\_SITE 201 201 BY SIMILARITY.  
 SQ SEQUENCE 735 AA; 82398 MW; 713104NMFABFBI CRG64;  
 Query Match 61.3%; Score 2489; DB 1; Length 735;  
 Best Local Similarity 64.2%; Pred. No. 9,8e-167;  
 Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;  
 QY 32 REMAGKGVWNGNT---STGTNKNWPGSLNDLILQDQRKSDPMDPNTREVR 88  
 DB 4 QNRQNAACP-FHGSVYNGSSNRTNKNWPNQLNLILQDQRKSDPMDPNTREVR 62  
 QY 89 KLDFALKKDLKALDKDSEWADWGHGGLMTRMWSAGTVTADRGGGGTGSORF 148  
 DB 6 KLDYNAKDLKLLKLESDWMDPAGYHGLPLTRMWSAGTVTADRGGGGTGSORF 148  
 QY 149 APLNSWPNNDLAKRLLPKIKKYKKNKSNADLMLAGTVVYSEGLPAGFSQVD 208  
 DB 6 APLNSWPNNDLAKRLLPKIKKYKKNKSNADLMLAGTVVYSEGLPAGFSQVD 208  
 QY 123 APLNSWPNNDLAKRLLPKIKKYKKNKSNADLMLAGTVVYSEGLPAGFSQVD 182  
 DB 6 APLNSWPNNDLAKRLLPKIKKYKKNKSNADLMLAGTVVYSEGLPAGFSQVD 182  
 QY 209 IWEPEKDYNGDEKEMIA--SERVSGD---RELENPLAAVQMGILVYNPESGDKPDK 267  
 DB 183 VHEPEEDYVNGSEKEMIA--SERVSGD---RELENPLAAVQMGILVYNPESGDKPDK 236  
 QY 268 RTAQOVLFTFANAMDEKFTALTAGHTVNGCHNGSNALAPPPASVENVQGLGN 327  
 DB 237 AARSADIREFRWGMDRETVALLAGCTFGKAGGAGATHVGEPEAPTEAQGLHS 296  
 QY 328 PMWGKASNAVSGEGAMTPTFKDMQYFDLLFTYHWEKKSPPAGHHPEDIDKEN 387  
 DB 297 STYGKSGSDTTSVSGEGAMTPTFKDMQYFDLLFTYHWEKKSPPAGHHPEDIDKEN 387  
 QY 388 KPVDAQSDSTHNPIDADWAKLVNVTYALCKEYHADPEYFKTKFANFKLHEDLG 447  
 DB 357 LAPDAEDSKVYPTNMATDILALRDEPEYKIAIRRFUNPEEFAEAFANFKLTHRG 416  
 QY 448 PRSRYGPEYPAEDLTWQDPTAGNTDYCE---EYVKTKAQSGLSISENVSTANDSNT 504  
 DB 417 PRTYVLAGPEYPAEDLTWQDPTAGNTDYCE---EYVKTKAQSGLSISENVSTANDSNT 504  
 QY 505 YRGSMDRGKAGARILAPQNEKNGENPERELAKVSYTEQISAD--TGASLADTVLAGS 562  
 DB 477 FNSDKRGKAGARILAPQNEKNGENPERELAKVSYTEQISAD--TGASLADTVLAGS 536  
 QY 563 VGEFAKAGADYRVVPLIGRGDARMTDASFAPIELADGFNRKQKVEYVVKPEM 622  
 DB 537 ANVEKAARDGADYRVVPLIGRGDARMTDASFAPIELADGFNRKQKVEYVVKPEM 622  
 QY 523 VADKALMGITGPEWVLGGMVLYNKGTVGKGVFTDCEGLTDFNLTDFNGSK 682  
 DB 597 LVDAKALMGITGPEWVLGGMVLYNKGTVGKGVFTDCEGLTDFNLTDFNGSK 656  
 QY 683 PVGSNAYIERDKGVAKTVASRVDFVFGSNLSRASAETAYQADNGKEKVFYDFAWTK 742  
 DB 657 PVGSNAYIERDKGVAKTVASRVDFVFGSNLSRASAETAYQADNGKEKVFYDFAWTK 716  
 QY 743 VNAADREDF 751  
 DB 717 VNAADREDF 725  
 RESULT 2  
 KX: 100%  
 KY: 100%  
 QY: 100%  
 DT: 100%  
 DT 15-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Peroxidase/catalase T (EC 1.11.1.6) (Catalase-peroxidase T).  
 OS Mycobacterium smegmatis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1772;  
 RN [1]  
 RP SEQUENCE FROM N.A. (NC01155).  
 RX MEDLINE=97021930; PUBMED=8870251.  
 RA Billman-Jacobs H, Sloan J, Coppel R.L.:  
 RT "Analysis of Isoniazid-resistant transposon mutants of Mycobacterium smegmatis";  
 RL FEMS Microbiol. Lett. 144:47-52(1996).  
 NN [2]  
 NC SEQUENCE FROM N.A.  
 RC S227;  
 RA Egler O, Talenti A.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND BROAD-SPECTRUM PEROXIDASE ACTIVITIES. MAY PLAY A ROLE IN THE INTRACELLULAR SURVIVAL OF MYCOBACTERIA.  
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.  
 CC -1- COFACTOR: binds 2 prothomene IX and 2 iron ions per tetramer.  
 CC -1- MISCELLANEOUS: DELETIONS OR DEFECTS IN KATG GENE CAUSE ISONIAZID (INH) RESISTANCE.  
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL PEROXIDASE/CATALASE SUBFAMILY.  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in Heidelberg. The EMBL outstation is a non-profit institution with no restrictions on its use by non-profit institutions as long as its content and its modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: X89718; CAA02566.1;  
 CC EMBL: U46841; CAA02575.1;  
 DR HSSP: P00431; 1BSK  
 DR InterPro: IPR00763; Bac\_Catase/prase.  
 DR InterPro: IPR002016; Peroxidase.  
 DR Pfam: PF00141; peroxidase; 1.  
 DR PRINTS: PRO0458; PROXIDASE.  
 DR TIGRUS: 19021930; HPT: 1.  
 DR PROSITE: P500435; PROXIDASE 2; 1.  
 DR PROSITE: P500436; PROXIDASE 2; 1.  
 DR PROSITE: P50873; PROXIDASE 4; 1.  
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.  
 FT ACT\_SITE 110 110 BY SIMILARITY.  
 FT ACT\_SITE 114 114 BY SIMILARITY.  
 FT METAL 277 277 IRON (HEME AXIAL LIGAND).  
 FT METAL 51 51 L -> M (IN REF. 2).  
 FT CONFLICT 374 374 L -> RT (IN REF. 2).  
 FT CONFLICT 374 374 DIFFERENT POSITIONS (REF. 2).  
 SQ SEQUENCE 739 AA; 81131 MW; A15A75ANJ5P85DA CRG64;  
 Query Match 54.6%; Score 2216.5; DB 1; Length 739;  
 Best Local Similarity 57.1%; Pred. No. 1.2e-147;  
 Matches 425; Conservative 98; Mismatches 198; Indels 23; Gaps 8;  
 QY 26 PRHRAIR-----ANSKGVWNG--GNFTGTGSKWPKPGSLNDLILQDQRKS 74  
 DB 2 PEDRTEDSPPIGEAOTDAPAGCPAGFGRIKPEPVAGSSRDWPNQLNLILQDQRV 61  
 QY 75 DPMDPNTREVRKLDLFDKVDVHALMTDSEWADWGHGGLMTRMWSAGTVTADRG 134  
 DB 62 NPLDEDFDYSVQNLJDYDVALRADIVFVWHTSDQMPADGPHYGLPFRMHAAGTV 121  
 QY 135 ADGRGGGCTGQRTAPLNSWPNNSDOKARLLPFIKKYKKNKSNADLMLAGTVV 194







OS Streptomyces coelicolor  
 CC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
 CC Streptomycineae; Streptomycetaceae; Streptomycetes;  
 OX NCBI\_TaxID=1902;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2) / M145;  
 RA MEDLINE=21596410; PubMed=12000953; Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Hargrett-Philby D., O'Neill M.A., Kuesel M.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor J.,  
 RA Rowood D.A., Lettice A., Woodward J., Farrell B.G., Parkhill J.,  
 RA "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RT Nature 417:141-147(2002).  
 RL  
 CC -1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND  
 CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 2 H2O2(O2) = O2(2) + 2 H2O.  
 CC -1- COFACTOR: Heme 2 and 2 iron ions per tetramer (By  
 CC similarity).  
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL  
 CC PEROXIDASE/CATALASE SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC The European Bioinformatics Institute. There are no restrictions on its  
 CC use, and the information contained here is available for commercial  
 CC modified and this statement is not removed usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL393106; CAB57412.1; -  
 DR ISSP; F00431; 14263; Bac. ctase/prase.  
 DR InterPro; IPR002016; Peroxidase.  
 DR Pfam; PF00141; peroxidase; 1.  
 DR PRINTS; PR00458; PEROXIDASE.  
 DR TIGRFams; TIGR00198; cat\_per\_HPI; 1.  
 DR PROSITE; PS00435; PEROXIDASE 1; 1.  
 DR PROSITE; PS00436; PEROXIDASE 2; 1.  
 DR PROSITE; PS00437; PEROXIDASE 3; 1.  
 DR OXfam; OXfam00000; Peroxidase; Iron; Heme; Hydrogen peroxide;  
 KW Complete proteome.  
 FT ACT\_SITE 105 105 BY SIMILARITY.  
 FT ACT\_SITE 109 109 BY SIMILARITY.  
 FT METAL 272 272 IRON (HEME AXIAL LIGAND).  
 FT SEQUENCE 740 AA; 80815 MW; 767C0A9BC9BD798 CRC64;  
 SO  
 QUARTY MATCH 53.0%; Score 2149.5; DB 1; Length 740;  
 Best Local Similarity 55.5%; DB 1; Length 740;  
 Matches 416; Conservative 111; Mismatches 192; Indels 31; Gaps 12;  
 OY 24 VSPHRAIRERAMSK---CPVMHG--GNTSTGTSKNDWPEGLMDLHQDQKSDPD 78  
 DB 1 MSNDIAIYDTAKTEEDGCPVHGRAPHTQGGNQRMPERLAKILKANVANPLG 60  
 OY 79 PPNPTEVRYKLDPAKQVIALMTDQSGWPPADKHGGLIMRAMHISAGTVIADGR 138  
 DB 61 EEFDYAEAPALDAAKVQDAIEVLTSDQWPDAGNGYPLMIAMHISAGTVIADGR 120  
 OY 139 GGGGTGSGFAPLNSWPDNVLKARLLMPLIKKYNKISWADLMLTLAGTVAESGLP 198  
 DB 121 GGAGGQGFAPLNSWPDNGLKARLLMPLIKKYNKISWADLMLTLAGTVAESGLP 180  
 OY 199 AVGSGFQGVIVWPKRIWGDKEWMLASDERY-GVNVKPTPMENPLAAMGLIYVP 257  
 DB 181 TGFAGAGVADWAEEDVYNGETTLW---DDRITGD---RELENPLQVAGGLIYVP 234

OY 258 EVNGHPPDLRTAQVLETFARMAWDEKTAALTAGHTGVCNCGNSALAPDPKASD 317  
 DB 235 EGPNNPDIIAARDINELTFARMAWDEKTAALTAGHTGFKTKGRPADAGDOPFAAA 294  
 OY 318 VENQGLGNGNPNQKASNAVTSCTEGAWTNTKPDNGYFDLLFGVNNELKSPAGAHH 377  
 DB 295 MEQGLGKMSHTGTCGGDAITSGLEVTWISTTSTQNGNFFKNLEFEETELESQAGANQ 354  
 OY 378 WPTDITKKNPV--DASDPSIRHNTDAMAKVNTYTPAICEKFMADPEVFKFYFA 435  
 DB 355 W-----VAKDAPELIDAPDAKAKHPRMLTIDSLRLDIYGPISRTERPEFADAPA 410  
 OY 436 KAWFKLTHRDGSPKSIKIGPEVPAEDLIWDQPI--PAGNTDYCEV--VKQIAQGLSI 491  
 DB 411 RAMEFLTHRDGSPKSIKIGPEVPEETLWQOPLPEPEGEVDAEDVATLKLLSGLSV 470  
 OY 432 SKMSTANDASAPYSGWGRGANGARTLALQNEHOGNRPRIKAVLSYE---QISA 547  
 DB 471 SOLVTTAWASATIFRSGDKGANGARLRLERFQGWNEVPEDELAQVLVLRVQRENS 530  
 OY 548 DTGA---STADVTLAGSVGEKAAKAGVYRVFVPLAGRGDAMTADSEAPULPLA 604  
 DB 531 GSGAKYVLAIDLVLGGSAKAAKEAGPEVPEFAGRVDAETHTDAESFEALPFA 590  
 OY 605 DGFENRQKVEYTKDPSKLSQALAGLTGTPKVIILGQNGYTGNTGTHWTTDCEG 664  
 DB 591 DGFENYLGKGRNLPFAEFLIDRLANLLTAPKMTVLGGVLVLAGHQSQSLGVFTPTG 550  
 OY 665 QLTNDNFVYALDNGSMKPVGSN--AYIHDRKTKAVKTAASRVOLVFGSNLSRYAEV 722  
 DB 651 SLTNDNFVYALDNGTTHKSTSEDTTFFGRDAATGEYKAGSRAOLVFGSNALRAEY 710  
 OY 723 QAUDDNGCFYVDFVAAATKVVNADRFVA 752  
 DB 711 YASDDAGECFVHDFVAAMKVKNLDREDA 740  
 RESULT 6  
 CATA\_ECOLI  
 OY P13029; ECOLI  
 AC P13029; STANAND; PRT; 726 AA.  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)  
 DE (Hydroperoxidase 1).  
 GN KATG KR 83942.11.  
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RA MEDLINE=88314956; PubMed=3045098;  
 RA "1998-Raine B.D., Dooley B.W., Mulvey M.R., Sorby P.A., Loeven P.C.;  
 RA coli" colli. colli. colli. colli. colli. colli. colli. colli. colli. colli.  
 RL J. Bacteriol. 170:4415-4419(1988).  
 RP  
 RP SEQUENCE OF 1-339 FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RC STRAIN-K12 / MG1655;  
 RA "Plunkett G.V., Burland V.D., Daniels D.L., Blattner F.R.;  
 RA coli" colli. colli. colli. colli. colli. colli. colli. colli. colli. colli.  
 RL Nucleic Acids Res. 21:3391-3398(1993).  
 RN  
 RP SEQUENCE OF 309-726 FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RC MEDLINE=94089392; PubMed=8265357;  
 RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,  
 RA Daniels D.L., Burland V.D., Plunkett G. III, Sofia H.J.,  
 RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the

	OY	360	LLEFYENNELKSPAGAHNHEPDIETKEKNKPVDASDSSTERNIPMDADMAIKYNPYRAI	419
D8	DB	332	NLFYEKWOTISFAGAIEFAEVD-APETID-PDFDSSKKRPMTLVLTDLTIRPDFESKI	389
	OY	420	CERFNADPEPKTKATAMFKLTHRDGPKSRITGEVPVPAEDTIWQDPITPAIGNTDCEEV	479
D8	DB	330	SRRFLNDPONFANRANKMLTHRNGPSKRITIGEVPVEDIIWDQPDPQPIYPNTBQD	449
	OY	480	V---KOKTAOSGLSISEMYSTDMSARTYGSGMRGGANGARILAFONHMGCSNEPERLA	536
D8	DB	450	IDULKFAJANDSLGSLSSELYSVANASTFTFGSDARGGANGARIALMPGDMDVN--AAAV	507
	OY	537	KVLVSVEEQIESADTG-ASTADVIVLAGSVTERKAAGAKOVDVRPFELRGGRDATMDDAD	595
D8	DB	508	RALPLFEIGREGBAAGGVVVYLAWVGEEKSAAGSLSHVFPGRVKARDQQPDTE	567
	OY	596	SENAPLEADLGPRNQNKKEYPVKDLDLRQAOLKGLETGPCTVLLGSRYVLGTNYGDK	655
D8	DB	568	MPELLPEPADGFNFARYARDVSYTESLLIDKAQOALTAEENTALLXGNRVYLANPGSK	627
	OY	656	HGVFDICDGQLTWDFNVNLTMGNSKPV-CGSNAIREORTKTAGVMRTASRVDAVFEGSN	713
D8	DB	628	NGTFDRVGYLSLIDFFVNLDMRKDTDESKELREGRETGEYKFSTRADRULVFEGSN	687
	OY	714	SLRLSTAEEVAQDNKEGVEPDVFAVTKVNNADDREV	751
D8	DB	688	SVLRBAEAYASSDAHREFVKDFVAAWKVNMILDREL	725
RESULT 7				
			CATA_SALTY	
			CATA_SALTY STANDARD; PRT;	726 AA.
			AC AC	P17750;
			01-FEB-1990 (Rel. 15, Created)	
			28-AUG-2003 (Rel. 41, Last sequence update)	
			15-SEP-2003 (Rel. 42, Last annotation update)	
			Peroxidase/catalase HPI (EC 1.11.1.6) [Catalase-peroxidase]	
			(Kof Oxydoreductase I).	
			KatG Oxydoreductase I). Salmonella typhimurium. Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella. NCBI_TextID:602; [ ]	
			SEQUENCE FROM N.A.	
			MEDLINE=U0111769; Pubmed=2277629;	
			Leowen P.C., Stauffer G.V.;	
			"Nucleotide sequence of katG of salmonella typhimurium L72 and characterization of its product, hydroperoxidase I"; Mol. Gen. Genet. 224:147-151(1990); [2]	
			SEQUENCE FROM N.A.	
			MEDLINE=D153496; PubMed=1677609;	
			McCllelland M., Sanderson K.E., Spieth J.J., Clifton S.W., Latreille P.-,	
			Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,	
			Leonard S., Nguyen C., Scott K., Holmes A., Grwal N., Mulvaney E.,	
			Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,	
			Waterson R., Wilson R.K.;	
			The complete genome sequence of Salmonella enterica serovar Typhimurium Nature 413:852-856(2001).	
			- FUNCTION: bifunctional, exhibiting both a catalase and broad-spectrum peroxidase activities.	
			- CATALYTIC ACTIVITY: 2 H(2)O2 -> O(2) + 2 H(2)O.	
			- COFACTOR: Binds 2 prothione IX and 2 iron ions per tetramer.	
			- SUBUNIT: Homotetramer.	
			- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL	
			- SIMILARITY: Contains a SHIMILAR regulatory motif (HRM) repeat.	







[illegible]



## RESULT 12



473	TDY	---CEWV--WOKIAQSLSISEVUSTAFDSARTFSGSGGNGANGARIRLAPQNE	526
474	DB		
475	DB		
446	VDY	KLVLDVANDZANLGRKINSLSUTTSSEIYKTAWASASTFTGTDGNGGANGARIRLAPQND	505
527	WGONE	PERLARIKLVSVYEQI---SAD-TGASIDAVTIAGSVGTGKAAAGYDVRY	578
506	IFPND	POELAKVIAKTDQSTFNNFNADQAGCKKISLADIVLGGHAAEQNAQKQGVDTIV	565
579	PELGR	KDQVATWTDQSTFNNFNADQAGCKKISLADIVLGGHAAEQNAQKQGVDTIV	565
579	PELGR	KDQVATWTDQSTFNNFNADQAGCKKISLADIVLGGHAAEQNAQKQGVDTIV	565
579	PELGR	KDQVATWTDQSTFNNFNADQAGCKKISLADIVLGGHAAEQNAQKQGVDTIV	565
566	PTF	PDRTDQATQGTQVQKSFPEVLEKKAQGFNTYFDKSNNSPDELVEKASLLKLSVPMT	625
639	VLLGSMV	YNTGNTGGTRHGVYTCDEGOITVDFVNTDGNKSKPVG--SNAYEIRDRKTI	696
639	VLLGSMV	YNTGNTGGTRHGVYTCDEGOITVDFVNTDGNKSKPVG--SNAYEIRDRKTI	696
636	VLGQSHR	KLVNANYQNGQYFDKPTQTNLDFNLLGSSPEWKKSSFTSEGTGTEYERKTI	685
697	GAVKV	TAWSRDLVPRGSSNLSLARSVYAAQDNGKQFVDFVAAPIKYNNADRFQV	751
686	GLKWK	ATSVLDLFGANSELARVAEAYATDQAEKFIQDPTNNAVYVMTADRFDI	740

RESULT 15

CATA\_HALMA STANDARD; PRF; 730 AA.

OS9651:

16-OCT-2001 (Rel. 40, Created)

28-FEB-2003 (Rel. 41, Last sequence update)

15-SEP-2003 (Rel. 42, Last annotation update)

Peroxidase/catalase (EC 1.1.1.6) (Catalase-peroxidase).

PERA. Cula marismortui (Halobacterium marismortui).

Achaeb. Euryarchaeota; Halobacteriia; Halobacteriales;

Halobacteriaceae; Haloarcula.

NCBI\_TaxId:2238; [1]

SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND MASS SPECTROMETRY.

MEDLINE=99122356; PubMed=9924978.

Cannaec-Caffrey Y., Hudry-Clergen G., Petitot Y., Gagnon J.,

"The protein sequence of an archaeal catalase-peroxidase.;"

Biochimie 80:1003-1011(1998).

-I- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND

BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).

-I- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.

-I- COFACTOR: Binds 2 prothemo IX and 2 iron ions per tetramer (By

MASS SPECTROMETRY: W0681292; W0688-9; METHUO-MALDI

-I- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL

PEROXIDASE/CATALASE SUBFAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

at the European Bioinformatics Institute. There are no restrictions on its

use, but the EMBL/EBI logo and the name of the source must be included.

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; Y16851; CAA76423.1; "

PIR: T48846; Y4846.

Trp; ITR; TRP0002.

InterPro: IPRO02036; Peroxidase.

Pfam: PF001141; peroxidase; 1.

PRINTS: PR00458; PEROXIDASE.

TIGRFam: TIGR010198; cat\_per\_HPI; 1.

PROSITE: PS00435; PEROXIDASE.1; 1.

PROSITE: PS00436; PEROXIDASE.2; 1.

PROSITE: PSS0873; PEROXIDASE.4; 1.

OXIDOREDUCTASE; Peroxidase; Iron; Heme; Hydrogen peroxide;

Structure.

INIT\_MET 0

ACT\_SITE 91 BY SIMILARITY.

```

CC CC ACCTOXYLASE; OXIDOREDUCTASE; HYDROLYTIC ACTIVITY;
CC CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).
CC CC -|- CATALYTIC ACTIVITY: 2 H2O(2) + O-2 + 2 H(2)O.
CC CC -|- COFACTOR: Binds 2 prothème IX and 2 iron ions per tetramer (By
CC CC similarity).
CC CC -|- MASS SPECTROMETRY: MW=81292; MW_ERR=9; METHOD=MALDI.
CC CC -|- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC CC PROTEIN; CATALASE SUBFAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed, changed by or for commercial
CC CC purposes. The full text of the disclaimer can be found at http://www.isb-sib.ch/announce/
CC CC or send an email to license@sib-ch.ch.
-----
CC CC EMBL; Y16851; CAAG76423.1; -.
CC CC PIR: T44846; T44846.
CC CC PDB: 1TK; 28-ANG-02.
CC CC InterPro: IPRO00763; Bac_ctase/prase.
CC CC InterPro: IPR002016; Peroxidase.
CC CC Pfam: PF001141; Catalase.
CC CC PRINTS: PR00458; PEROXIDASE.
CC CC TIGRFAM: TIGR00198; cat_per_HPI_1.
CC CC PROSITE: PS00435; PEROXIDASE_1; 1.
CC CC PROSITE: PS00436; PEROXIDASE_2; 1.
CC CC PROSITE: PS00873; PEROXIDASE_4; 1.
CC CC Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
CC CC catalytic activity.
CC CC INIT_MET 0
CC CC ACT_SITE 91 BY SIMILARITY.
CC CC FT

```

```

FT ACT_SITE 95 95 BY SIMILARITY.
FT RETAIL 256 256 IRON (HERE AXIAL LGAND).
SQ SEQUENCE 730 AA; 81252 MW; 5C7FFEC637A0C45 CRC64;

Query Match 49.2%; Score 1998; DB 1; Length 730;
Beat Local Similarity 52.7%; Pred. No. 2.5e-132;
Matches 385; Conservative 103; Mismatches 219; Indels 24; Gaps 8;

QY 36 MSKGCPCVHGGHNTSTGSKNDMPGJLNDILAQDQRKSDMPDFTNREVRKLDLFDAL 95
DB 8 MSKGT---CGSRKRPKSNOMWPSKUNLSTLDONARDVGVDDPDFAEEFKLIDKAV 63
QY 96 KKDVIHALTDSQENHWDKMGHGGIMTMAHHSAGTYRTADRGGGGTGSRFAPLNSWP 155
DB 64 KSDLELMTSSODWNPADYCHYCPLEPTMAHHSAGTYRTADRGGAAGGRFAPINSWP 123
QY 156 DNVSLQKARLLMPTIKKYNKISMDLMLTACTVAVESMGLPAYCFSGRVDIWEPEKD 215
DB 124 DNANLQKARLLMPTIKKYNKISMDLMLTACTVAVESMGLPAYCFSGRVDIWEPEKD 183
QY 216 IYNGDEKELAPSDERYGVNKPSTENPDLAANGVLITVNPESVNGHPDLRTAQVLE 275
DB 184 IYNGDEPEP--ETQERF--DEPGETQGLGAGVGLITVNPESGPDGNPDPEASAKNIQ 238
QY 276 TFRANNDKETAALFAGHTVNGHNGHNSA--LAPDFASDVENQGLGNGNPMQGRKA 334
DB 239 TFRANNDKETAALFAGHTVNGHNGHNSA--LAPDFASDVENQGLGNGNPMQGRKA 298
QY 335 SNAVTSIGEGAMTNTFKDMGYFDLLFCYNWELKSPACAHWEPIDIKKENKPYDASD 394
DB 299 GEMTISGIEGPMPTSTMDMGYINLLDYEMEPKGGGAWHAPKSELRKSVDPARD 358
QY 395 PSTRNPTMDADMAIKVNPYTHALCEKPMADPEYFKTKAKWFLTHRDLGPKSYIG 454
DB 359 PDEKOTPMULTTDALKRDPYREVNETQENMPEFCMFKAKNYLLTHDHCPPERFLG 418
QY 455 PEVPAEDLIWQDPTIAGNTDYC--EYV--VKQIAQSGLSISEMYSTARDASRTYRSDM 510
DB 419 PEVPEDEMIWQDPLDADYDLDGDEIAELKEBILDSLSVQLVKTAMASASTYRSDK 478
QY 511 RGGANGARIRLAPONKQNEPERLAKVLYTRQISAD-----TCASIAQVYLAGS 562
DB 479 RGGANGARIRLAPONKQNEPERLAKVLYTRQISAD-----TCASIAQVYLAGS 538
QY 563 VGTAKAAAGYDVRVPTLKGKGDATKEMTDADSFAPLEPLADGFRNMOKKEYVVKPEEM 622
DB 539 VGTAKAAAGYDVRVPTLKGKGDATKEMTDADSFAPLEPLADGFRNMOKKEYVVKPEEM 598
QY 623 LLDLQALMLGTCPEMTVLGGNRVLYGNTYGHGTHGYTCBGLTNDQFTNLTDMGNSWK 682
DB 599 LYVNDALILNITASELITAGISRSIGANYQDTDLGVFTDPEPTLTNDFFVNLDMGTWE 658
QY 693 PVGSNAYEIR--DKRTGAWKATGRVDLVFGSNNLSRSAYEYVAQDDNGKEKVFDFVAAW 740
DB 659 PADSEHRYKGLWDRDGEVKNWATRIDILFGSDNRLRAISEYVGSADAERKLVHDFVDTW 718
QY 741 TKYNAAGRDGV 751
DB 719 SKVVKLDRFDL 729

```

Search completed: October 7, 2003, 19:43:00  
Job time : 20.5861 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 06:25:58 ; Search time 8206.54 Seconds  
(without alignment)  
11276.082 Million cell updates/sec

Title: US-09-884-889-5

Perfect score: 2262

Sequence: 1 atgaacacatcgctgta.....gttcgacgtcgctgtaa 2262

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

1: gb-pat.\*

2: gb-htg.\*

3: gb-htg.\*

4: gb-em.\*

5: gb-em.\*

6: gb-pat.\*

7: gb-ph.\*

8: gb-pl.\*

9: gb-pr.\*

10: gb-pr.\*

11: gb-sts.\*

12: gb-sts.\*

13: gb-un.\*

14: gb-vi.\*

15: em-ba.\*

16: em-ba.\*

17: em-ba.\*

18: em-ba.\*

19: em-mu.\*

20: em-or.\*

21: em-or.\*

22: em-or.\*

23: em-or.\*

24: em-ph.\*

25: em-ph.\*

26: em-ro.\*

27: em-sts.\*

28: em-un.\*

29: em-vi.\*

30: em-htg-hum.\*

31: em-htg-inv.\*

32: em-htg-other.\*

33: em-htg-mus.\*

34: em-htg-mus.\*

35: em-htg-rod.\*

36: em-htg-man.\*

37: em-htg-vrt.\*

38: em-sy.\*

39: em-htgo-hum.\*

40: em-htgo-mus.\*

41: em-htgo-other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2262	100.0	2462	6	AR098263 Sequence
2	2262	100.0	2462	6	AR098263 Sequence
3	1198	50.8	14492	1	AR0404233 Sequence
4	1098	48.5	7772	1	AR0082233 Agrobacte
5	1098	48.5	10299	1	AR009392 Agrobacte
6	1098.4	48.5	2640	1	AR033631 Agrobacte
7	1085	48.0	296100	1	AF050085 Vibrio pa
8	1076	47.6	2700	1	AF197161 Synchococ
9	1073	47.4	304450	1	AF016806 Vibrio vu
10	1071	47.4	304450	1	AF016806 Vibrio vu
11	1030.4	45.6	3865	1	AF078110 Lactoba
12	1019.2	45.1	2430	1	SYOGP
13	951.2	42.1	12293	1	AE015873
14	945.2	41.8	2896	1	SYCKATG
15	945.2	41.8	146174	1	D90910
16	837.2	37.0	2208	1	AB020234
17	837.2	37.0	2208	1	AB020065
18	837.2	37.0	2232	1	AB020067
19	837.2	37.0	2235	1	AB020067
20	837.2	37.0	2235	1	AB020092
21	837.2	37.0	2238	1	AB020115
22	837.2	37.0	2241	1	AB020094
23	837.2	37.0	2247	1	AB020078
24	837.2	37.0	2250	1	AB020106
25	837.2	37.0	2251	1	AB020119
26	837.2	37.0	2253	1	AB020119
27	837.2	37.0	2256	1	AB020064
28	837.2	37.0	2256	1	AB020066
29	837.2	37.0	2256	1	AB020071
30	837.2	37.0	2256	1	AB020072
31	837.2	37.0	2256	1	AB020073
32	837.2	37.0	2256	1	AB020074
33	837.2	37.0	2256	1	AB020075
34	837.2	37.0	2256	1	AB020076
35	837.2	37.0	2256	1	AB020077
36	837.2	37.0	2256	1	AB020079
37	837.2	37.0	2256	1	AB020080
38	837.2	37.0	2256	1	AB020081
39	837.2	37.0	2256	1	AB020082
40	837.2	37.0	2256	1	AB020083
41	837.2	37.0	2256	1	AB020084
42	837.2	37.0	2256	1	AB020085
43	837.2	37.0	2256	1	AB020086
44	837.2	37.0	2256	1	AB020087
45	837.2	37.0	2256	1	AB020088

## ALIGNMENTS

RESULT 1	AR098263	Sequence	5 from patent US 6074860.	2262 bp	DNA	linear	PAT 14-FEB-2001
LOCUS	AR098263	Sequence	5 from patent US 6074860.	2262 bp	DNA	linear	PAT 14-FEB-2001
DEFINITION	AR098263	Sequence	5 from patent US 6074860.	2262 bp	DNA	linear	PAT 14-FEB-2001
ACCESSION	AR098263	Sequence	5 from patent US 6074860.	2262 bp	DNA	linear	PAT 14-FEB-2001
KEYWORDS	AR098263.1	GT:12807520					
SOURCE	Unknown.						
ORGANISM	Unknown.						
REFERENCE	1 (bases 1 to 2262)						
AUTHORS	Robertson,D.E., Sanyal,I. and Adhikary,R.S.						
TITLE	Catalases						
JOURNAL	Patent: US 6074860-A 5 13-JUN-2000;						
FEATURES	Location/Qualifiers						

source	1. .2262	Query Match	100.0%	Score 2262;	DB 6;	Length 2262;			
BASIS COUNT	553 a 612 c 663 g 434 t	Best Local Similarity 100.0%; Pred. No. 0;							
ORIGIN		Matches 2262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
Qy	1	ATGATATACGATCCGCTGACGATCTACACAGTACGTTGACGAAAGATGACAGCATTTT	60						
Db	1	ATGATATACGATCCGCTGACGATCTACACAGTACGTTGACGAAAGATGACAGCATTTT	60						
Qy	61	GTGCTCTTGGTATGCGCAAGCATAGACAAATAGGGAAGATATGACGGGTAATGT	120						
Db	61	GTGCTCTTGGTATGCGCAAGCATAGACAAATAGGGAAGATATGACGGGTAATGT	120						
Qy	121	CTGTCATGACGGGTGATACAGCTGACCGGTACTGCAACAAAGATGTTGGTGGCGAA	180						
Db	121	CTGTCATGACGGGTGATACAGCTGACCGGTACTGCAACAAAGATGTTGGTGGCGAA	180						
Qy	181	GGGTGAACCTGGATATTTGGCATCAGCAAGATCGCAAAATCAGACCGATGGATCGGAT	240						
Db	181	GGGTGAACCTGGATATTTGGCATCAGCAAGATCGCAAAATCAGACCGATGGATCGGAT	240						
Qy	241	TTGACATACCGGTGATGATGACGATGATGATGATGATGATGATGATGATGATGATG	300						
Db	241	TTGACATACCGGTGATGATGACGATGATGATGATGATGATGATGATGATGATGATG	300						
Qy	301	GGGTGATGACCGATAGCAAGATGTTGGCGCGCTGACTGGGGCAGCTACGGCGGTTTG	360						
Db	301	GGGTGATGACCGATAGCAAGATGTTGGCGCGCTGACTGGGGCAGCTACGGCGGTTTG	360						
Qy	361	ATGATCGGTATGCGCTTGGCACTGCGGTGGGCACTACCGTATGCTGATGCTGATGCTG	420						
Db	361	ATGATCGGTATGCGCTTGGCACTGCGGTGGGCACTACCGTATGCTGATGCTGATGCTG	420						
Qy	421	GGTGGTACCGGAACGCGCTTTGGACCGCTCACTCTGCGCGGACACAGCTCAGCGTG	480						
Db	421	GGTGGTACCGGAACGCGCTTTGGACCGCTCACTCTGCGCGGACACAGCTCAGCGTG	480						
Qy	481	GATTAAGCGCGCTGCTGCTGGCGCATCAAGAGAGTACGGCAACAAATCAGCTGG	540						
Db	481	GATTAAGCGCGCTGCTGCTGGCGCATCAAGAGAGTACGGCAACAAATCAGCTGG	540						
Qy	541	CGACAGCTGATGCTTGGCTGGCAACGCGGTGATGCTGATGCTGATGCTGATGCTG	600						
Db	541	CGACAGCTGATGCTTGGCTGGCAACGCGGTGATGCTGATGCTGATGCTGATGCTG	600						
Qy	601	GGCTTCTCTTTCGGCGCGGTGATATTTGGGAACCCGAAAGATATCTACTGGGTGAC	660						
Db	601	GGCTTCTCTTTCGGCGCGGTGATATTTGGGAACCCGAAAGATATCTACTGGGTGAC	660						
Qy	661	GAAGAGAGTGTGCTGACCTCTGACGACGCTACGGCGACGTTGACACCGCAGAGACC	720						
Db	661	GAAGAGAGTGTGCTGACCTCTGACGACGCTACGGCGACGTTGACACCGCAGAGACC	720						
Qy	721	ATGGAAGAGTGTGCTGACCTCTGACGACGCTACGGCGACGTTGACACCGCAGAGACC	780						
Db	721	ATGGAAGAGTGTGCTGACCTCTGACGACGCTACGGCGACGTTGACACCGCAGAGACC	780						
Qy	781	ACGGCAACCTGATCCGCTGAGACGCGACAGAGGTACTTGAAACCTTTCGCCCGTATG	840						
Db	781	ACGGCAACCTGATCCGCTGAGACGCGACAGAGGTACTTGAAACCTTTCGCCCGTATG	840						
Qy	841	CGCATGAAAGAGTGTGCTGACCTCTGACGACGCTACGGCGACGTTGACACCGCAGAGACC	900						
Db	841	CGCATGAAAGAGTGTGCTGACCTCTGACGACGCTACGGCGACGTTGACACCGCAGAGACC	900						
Qy	901	CAGGTATGAGTATGCTTGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG	960						
Db	901	CAGGTATGAGTATGCTTGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG	960						

[illegible]

[illegible]





Db 10538 TACTGGGGTTCTGAAAGATGGCTTGCCAAATCTGGCGGTGAAATAGCGTTACTCT 10597  
 QY 709 AAGCCAGAGACATGAGAAAGCGGCTCGGGCGCTGCCAATGGGCTGATGCTATGTGAAC 768  
 Db 10598 GGTACGCCGATCTGAAAGCCCATTTGCGGCGTCATGATGGGGCTTATCTAGCTACAC 10657  
 QY 769 CCGGAAGGCTGTAAACGGCCACCCCTGATCGGTGAGACCGCAGACAGGTACTTGAAAC 828  
 Db 10658 CCAAGAGGTGGATGGAACACCGAGATCCACTCAAAAGCGGCAAGACATGGGTGCAT 10717  
 QY 829 TTGGCGCTGATGGGATGAGACAGCAAAAGCCAGCGGCTACACTGTGGGCGACAC 888  
 Db 10718 TTTCGCTGCATGGCAATGAATGACGAGAGACGGTTGCTCTGACGCTGGCGGCATACG 10777  
 QY 889 GTGGGTAATTTGATCGCAATGCGCAATGCGTCTGTGGCTTAGCCCTCGAACCAAGCGCT 948  
 Db 10778 GTGGGTGAAGCGACAGGTATGAGCAAAAGCTTCTACCTCTGGCGCTGATCGAGGGCT 10937  
 QY 949 GAGGTTGAAGCAAGGCTTAGGTTGGGCAACCCACACGACGTCGAGCGGCGAGAGG 10987  
 Db 10938 GAATGCAACCAACAGGCTTAGGTTGGAAACACACACAGCGCTGGGATTTGTCGAAC 10987  
 QY 1009 CCGGTGACCTCGGGTATGGAAGATGCTTGGAACACCAACCCACGAGAAATCGATATGGCG 1068  
 Db 10988 ACCTGACCAAGTGTATGAGGTGTCATGACACCTCAACCAACCCGCTGGGATGAGAA 10957  
 QY 1069 TATTTCGACCTGCTTGGGCTACAAATGGGAATGAAAGAGCTCTGCCGCTGGCCAC 1128  
 Db 10958 TATTTCCTACTTACTGCTCAGCTATGCAATGGCAGCTCACTAAGAGCCCTCGAGTGCATGG 11017  
 QY 1128 CATTGGGACCGATGACATCAAAAGGAAACAGCGGTTGACGCCACGACCCCTCT 1188  
 Db 11018 CATGGGAACTGCTCACTACAAAGAGATTAACCGGTGATGTGGAAGCAACGAT 11077  
 QY 1189 ATTGSCACACCGCATGACGATGCGATGCGATATGCGATAAAGAGTAAATCCGACCTAT 1248  
 Db 11078 ATTGCTACACACCATGATGATGACCGCATGCGCATATGTCCTCAAAATCGATCTGAATAT 11137  
 QY 1249 CGCGCTATCTGGGAAATTCATGCGCGCATCTGAGTACTTCAAGAAACACTTTGGCGAG 1308  
 Db 11138 GCGAGATTCAGAGCGCTCTCAAAAGATCCCGTACTCTCAAGATGTTGCGCA 11197  
 QY 1309 SCGTGTTTCAAGCTGACGACGCTGACCTGGGCGGAAATACGTTACATCGGCCCGGAA 1368  
 Db 11198 CTTTGGTTAAAGCTCACTACCGCATATGAGGACAAAGCCGCTACTTTGGCCCATG 11257  
 QY 1369 GTGCGGCGAGACGATGATTGGGAAGACCCGATTCGGCAGGTACACCGACTACTGC 1428  
 Db 11258 GTTCGGGCAAGATTAATGCTGGCAAGACCTGTTCCAGAGCTTGAAGTATGAT 11317  
 QY 1429 GAAGAAGTGTCTACAGCAAAATTCAGAAAGTGGCTGAGCATTAAGTATGATGCTCC 1488  
 Db 11318 GTCAATGGCTCAAGCAAAATTCGCGCAAGCGTCTTAAGCATCATGTAAGTATGGTCAGC 11377  
 QY 1489 ACCTGTGGGACATGCTCGCTATTTATGCGGTTTCGATATGCGCGCGCTGTACMGCT 1548  
 Db 11378 ACCTGTGGGATACCTTCGCTACTTTCCTGTTCTGTAAGCTATGGGAGGACAGG 11437  
 QY 1549 GCGCGCATCGTTGGCCACAGACAGGATGGGAGGCGACACCGGAGCGGCTGGCG 1608  
 Db 11438 CCGCGTATTCGCTCTGCTCACAAAGATTTGGGAAGCAACGACTCGCGGCTTGTG 11497  
 QY 1609 AAGTGTGAGCGTCTACGACAGATCTGTGGCGACACCGCGCTAGCATCGCGGACG 1668  
 Db 11498 AAGTCTGTGGATTTAGAGAAATTTGCCCTGAACTTGGCATGATGATGATGATG 11557  
 QY 1669 ATGTTCTTGGCGGATAGCTGATGAGCAATCGAAGACCGGCAAGACGATGATGATG 1728  
 Db 11558 ATGTTCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 11617  
 QY 1729 CGGTTTCCTCTCTGAAAGCGGCGGAGATGCGGAGATGCGGAGATGACCGGACGAGCTCC 1788  
 Db 11618 ACCTGACGCTTGGCACCAGGAGCGGCGATGCGGCAATGCGAAGACGATTTGAGTGC 11677

QY 1789 TTGCGACCGCTGGAGCGCGTGGCGGATGCTTCCCAACTGGCAGAGAAGAAGATGTGTG 1848  
 Db 11678 TTGAGATGCTTGAAGCTCTCGCGGATGCTTCCGTAACGTCAGAAAGCAATATGTGTG 11737  
 QY 1849 GTGAGCGCGGAGAGATGCTGCTGGATGCGGCGAGCTGATGGCTTAACCGGCGCGGAA 1908  
 Db 11738 GTACACCGGAGAAATGTTCTGCGACAAGGCACTGCTGCTCAGCGCTCCAGAA 11797  
 QY 1909 ATGACCGTGTCTGCTGGCGGTATGCGGCTACTGGCGCACTATGCTGGCGCAACAC 1968  
 Db 11798 ATGACGCGTATGTTGGTGGCATGCTGCTGTGCGCACTACTAGTGGGCGACCAAC 11857  
 QY 1969 GCGGTATTCAACGATTTGGAGGCGGCTTGACCAACGACTTTTGTGCACTGACGAT 2028  
 Db 11858 GCGGTGTTTACCGACCGCGTGGTGGCTGTGACCAACGATTTCTGCTTCACTTCAACGAC 2087  
 QY 2029 ATGGGACACGCTGGAGACCGGTAGGTAGCAACGCTCTAGAAATCGCGCGACGACGAC 2088  
 Db 11918 ATGATGTACACTGGAAACACTGCTGCACTCTTACGATTTGAAAGTAAAGC 11977  
 QY 2089 GTGCGCTGTAATGTGACCGCTCGCGGCGGTGATGCTGTTATTTGGTTCACACTGCGTACTG 2148  
 Db 11978 GCGAAATGAAATGGACAGCACACTGCTGTGATCTGGTGTGTTGTTCTACTCATCTCA 12037  
 QY 2149 GCTCTTACCGAGAGTGTAGCGCCAGGACGACGACGACGACGACGACGACGACGACGACG 2208  
 Db 12038 GTGCTCTATCGGAGTGTAGCGCAGGACGACGACGACGACGACGACGACGACGACGACG 12097  
 QY 2209 GTGCGCGCTGGACCAAGTGTAGGCGGACGCTTTTCGACGTGCGGT 2257  
 Db 12098 GTTGGCGCATGGACCAAGTAAATGATGCGGATCGATTTGATCTGTGT 12146

RESULT 4  
 AE008223/c  
 LOCUS Agrobacterium tumefaciens str. C58 linear chromosome, section 27 of 187 of the complete sequence.  
 ACCESSION AE008223 AE007870  
 VERSION AE008223.1  
 SOURCE Agrobacterium tumefaciens str. C58 (Cereon)  
 ORGANISM Agrobacterium tumefaciens str. C58 (Cereon)  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 REFERENCE 1 (bases 1 to 7772) Hinkle, G., Slater, S.C. and Goodner, B. Characterization of Agrobacterium tumefaciens C58 genome. Sequences of Agrobacterium tumefaciens C58 (strain=C58), the Cauloviral Agent of Crown Gall Disease in Plants  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 7772) Hinkle, G., Slater, S.C. and Goodner, B. Direct Submission  
 JOURNAL Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA  
 COMMENT Agrobacterium tumefaciens C58 genome sequence missing from the left end of the chromosome and 200 bp missing from the right end.  
 FEATURES  
 Location/Qualifiers  
 1. 7772  
 /organism="Agrobacterium tumefaciens str. C58 (Cereon)"  
 /mol\_type="genomic DNA"  
 /strain="C58"  
 93..962  
 /taxon="181661"  
 93..962  
 /gene="AGR\_L\_475"  
 93..962  
 /note="AGR\_L\_475"  
 /note="(AF302797) Pobr"  
 /codon\_start=1  
 /transl\_table=11  
 /product="AGR\_L\_475p"





## AUTHORS

Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E., Chou,Y., Kozak,M., Kijima,M., Almeida Jr.,R.F., Zhou,Y., Boyes Sr,D., Chapman,P., Chaudhry,A., Schenck,J.G., Gillet,W., Grant,C., Guenther,D., Kutyavin,T., Levy,R., Li,R., McLelland,E., Palmeri,A., Raymond,C., Rouse,G., Saenphimachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I., Karp,P., Romero,P., Zhang,S., Too,H., Tao,Y., Biddle,P., Jung,M., Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C., Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V., et al.

## TITLE

Direct Submission

## JOURNAL

Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA

## FEATURES

source Location/Qualifiers  
1. .10299  
organism="Agrobacterium tumefaciens str. C58 (U.  
10299)  
mol\_type="genomic DNA"  
/strain="C58"  
/db\_xref="taxon:180935"  
121. .795  
/gene="Cops"  
/note="synonym: Atu4638"  
121. .795  
/gene="Cops"

## gene

121. .795  
/note="synonym: Atu4638"  
121. .795  
/gene="Cops"

## CDS

121. .795  
/note="synonym: Atu4638"  
121. .795  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/product="two component response regulator"  
/protein\_id="AAU45432.1"  
/db\_xref="GI:17743135"

## gene

121. .795  
/note="synonym: Atu4638"  
121. .795  
/gene="Cops"

## CDS

121. .795  
/note="synonym: Atu4638"  
121. .795  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/product="two component response regulator"  
/protein\_id="AAU45432.1"  
/db\_xref="GI:17743135"

## gene

121. .795  
/note="synonym: Atu4638"  
121. .795  
/gene="Cops"

## CDS

121. .795  
/note="synonym: Atu4638"  
121. .795  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/product="two component response regulator"  
/protein\_id="AAU45432.1"  
/db\_xref="GI:17743135"

## gene

121. .795  
/note="synonym: Atu4638"  
121. .795  
/gene="Cops"

## CDS

121. .795  
/note="synonym: Atu4638"  
121. .795  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/product="two component response regulator"  
/protein\_id="AAU45432.1"  
/db\_xref="GI:17743135"

## gene

121. .795  
/note="synonym: Atu4638"  
121. .795  
/gene="Cops"

## CDS

121. .795  
/note="synonym: Atu4638"  
121. .795  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/product="two component response regulator"  
/protein\_id="AAU45432.1"  
/db\_xref="GI:17743135"

## gene

121. .795  
/note="synonym: Atu4638"  
121. .795  
/gene="Cops"

## CDS

121. .795  
/note="synonym: Atu4638"  
121. .795  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/product="two component response regulator"  
/protein\_id="AAU45432.1"  
/db\_xref="GI:17743135"

## gene

121. .795  
/note="synonym: Atu4638"  
121. .795  
/gene="Cops"

## CDS

121. .795  
/note="synonym: Atu4638"  
121. .795  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/product="two component response regulator"  
/protein\_id="AAU45432.1"  
/db\_xref="GI:17743135"

## gene

PKYVEDAYGRSGLOAGVADKFGAPENTDKHAGSR\*  
complement(3171. .4079)  
/gene="OxyR"  
/note="synonym: Atu4641"  
complement(3171. .4079)  
/gene="OxyR"

## CDS

141. .4079  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/trans\_table=11  
/product="transcriptional regulator, LysR family"  
/protein\_id="AAC45435.1"  
/db\_xref="GI:17743137"

## gene

141. .4079  
/note="synonym: Atu4642"  
141. .4079  
/gene="kita"

## CDS

141. .4079  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/trans\_table=11  
/product="catalase"  
/protein\_id="AAL45436.1"  
/db\_xref="GI:17743138"

## gene

141. .4079  
/note="synonym: Atu4643"  
141. .4079  
/gene="kita"

## CDS

141. .4079  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/trans\_table=11  
/product="catalase"  
/protein\_id="AAL45436.1"  
/db\_xref="GI:17743138"

## gene

141. .4079  
/note="synonym: Atu4644"  
141. .4079  
/gene="kita"

## CDS

141. .4079  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/trans\_table=11  
/product="catalase"  
/protein\_id="AAL45436.1"  
/db\_xref="GI:17743138"

## gene

141. .4079  
/note="synonym: Atu4645"  
141. .4079  
/gene="kita"

## CDS

141. .4079  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/trans\_table=11  
/product="catalase"  
/protein\_id="AAL45436.1"  
/db\_xref="GI:17743138"

## gene

141. .4079  
/note="synonym: Atu4646"  
141. .4079  
/gene="kita"

## CDS

141. .4079  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/trans\_table=11  
/product="catalase"  
/protein\_id="AAL45436.1"  
/db\_xref="GI:17743138"

## gene

141. .4079  
/note="synonym: Atu4647"  
141. .4079  
/gene="kita"

## CDS

141. .4079  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/trans\_table=11  
/product="catalase"  
/protein\_id="AAL45436.1"  
/db\_xref="GI:17743138"

## gene

141. .4079  
/note="synonym: Atu4648"  
141. .4079  
/gene="kita"

## CDS

141. .4079  
/note="Identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/trans\_table=11  
/product="catalase"  
/protein\_id="AAL45436.1"  
/db\_xref="GI:17743138"



Db	5992	GGCGCGCTTGAGCGCGCTGGCGGATGTTTTCGGCACTGGGTTGAGAGAGCACTAATGTGCGTC	6051
QY	1852	AAGCCGGAAGAGATGCTCTGGATGCTGCGAGCTGATGGGCTTAACCGCCGCGGAAATC	1911
Db	6052	AGCCCGGAAGAGCTGCTCTGGATGCTGGCGACAGCTTTTGGGCGCTCACCGCGCGCAACTC	6111
QY	1912	ACGTGCTCTGCGCGCGGTATGCGGGTACTGGGACCAACTATGGTGGCGACCAACAACCGGC	1971
Db	6112	GATTCCTCATGTCGCGGGTCTCGGGGTCTATGGGCGCAATTAAGCGGTGGCGCGATGGC	6171
QY	1972	GTATTCACCGATTTCTGAAGCGCAGTGTGAACAAGCACTTTTGTGGTCTGAACCTGACCGAATG	2031
Db	6172	GTCTCTACCGATTAAGCGGGGGCGCTTTACAGAGCACTTTCTTCAGGAGCTGTGACGGACATC	6231
QY	2032	GGGACAGCTGGAACCGCGGTAGCTAGCAACGGCTACCAATCCGATACGCGACCGACAGACCGGT	2091
Db	6232	GCTATCTCTGGTCCGCGACCGCAACAATCTATAGCATCGGTGATGCGACGCGCGC	6291
QY	2092	CCGCTGAAGTGGACCGCTCTCGGGGTGGATCTGGTATTTTGGTCTCAACTCTGCTATCGCGC	2151
Db	6292	GCTACCGAGATATTCGCAACCGCGCGTCTGATCTGGTACGGCTCCCACTCATCTCTCGGC	6351
QY	2152	CTGACCGAGAGTGTACCGCCGCGCAAGTACGCGCGAGATGGTGTCAAGAGATGTC	2211
Db	6352	GCTATTGCGGAAGTTATGCGCAGGACGACACAGGGAATATTCGCGCGACTTCATT	6411
QY	2212	CCGCGCTGGACCAAGTCGATGAAGCGCGACGCTTTTCGA	2249
Db	6412	CCGCGCTGGACGCGAAGTCGATGAAGCGCGACGCTTTTCGA	6449

RESULT 6	
AB033631	
LOCUS	2640 bp DNA linear BCT 09-FEB-2000
DEFINITION	Agrobacterium tumefaciens kta gene for catalase, complete cds.
ACCESSION	AB033631
VERSION	AB033631.1 GI:6691547
KEYWORDS	Agrobacterium tumefaciens (Rhizobium radiobacter)
SOURCE	Agrobacterium tumefaciens
ORGANISM	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
REFERENCE	Xu,X.Q. and Pan,S.Q.
AUTHORS	1 (sites)
TITLE	Agrobacterium catalase is a virulence factor involved in tumorigenesis
JOURNAL	Mol. Microbiol. 35 (2), 407-414 (2000)
MEDLINE	20118000
PMID	10652101
PUBLISHED	10652101
REFERENCE	2 (bases 1 to 2640)
AUTHORS	Xu,X.Q. and Pan,S.Q.
TITLE	Direct Submission
JOURNAL	Agrobacterium tumefaciens (Rhizobium radiobacter)
COMMENT	117543, Singapore (E-mail: sculp31@nus.edu.sg, Tel:636743596)
FEATURES	Sequence updated (23-OCT-1999), Location/Qualifiers
source	1. 2640
gene	/organism="Agrobacterium tumefaciens"
	/mol_type="genomic DNA"
	/db_xref="taxon:358"
	308..2479
	/gene="kta"
CDS	/gene="kta"
	308..2479
	/function="hydrogen peroxide detoxification"
	/codon_start=1
	/transl_table=1
	/product="catalase"
	308..2479
	/db_xref="GI:6691548"
	/translation="MDATSKPAKGVNHHGNTVSGKSVTEWPNALNLIIDHDTK
	TNPAGTSFNREALTQIDVLRALMTDRLALMTDQEWDPAGVYVGMWARYTHAAGS

Query Match	48.5%	Score 1096.4	DB 1	Length 2640
Best Local Similarity	69.6%	Pred. No. 4.1e-219	Indels	Gaps 0
Matches 1407	Conservative	0	Mismatches 631	
QY	112	GGTAAATGCTGTGTACATGACAGGTGGTGTACATGCTGACAGCGGTACTTCCACAAAGATGG	171	
Db	332	GCGCAAGTGTCCGTCATGCAATGAGAGCAATACGGTCTCCGCGCAAAATCGGTGACCGAATGG	391	
QY	172	TGGCGGAGAGGGTGAACCTGGATATTTGGATCAGCAAGATCGCAAAATCAGACCCGATG	231	
Db	392	TGACAGAGGGTGAACCTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG	451	
QY	232	GATCCGGATTTCAACTACCTGCGAAGAGTACGCAAGCTCGATTTGACGCGGTGACGCGTGAAGA	291	
Db	452	GCGCACTCTCTTCAACTACCTGCGAAGAGCTGAGACGCGTTGATGTGGAAGCGCTTCAAGGCC	511	
QY	292	GATGTCACGCGTTGATGACCGATGACGCAAGATGTGGCGCGCTGACTGGGGCGACATC	351	
Db	512	GATGTCGCGCGCTTATGACCGCAATGCAAGATATGGTGGCGCGCGACATGCGCGCGCTTAT	571	
QY	352	CGCGGTTTGATGATCGCTGTGCGACCTCCGTCGCGACCTACCGTATGTCTGATGCGC	411	
Db	572	GTGCGCATGATGGCCGCTGTACTCGCATGCGCGCGGTTCTTATGCTGTGTCAGACAGCT	631	
QY	412	GTGGCGCGGTGTGATCGGAGACGACGCTTTGACCGCTCACTCTGCGCGGACAC	471	
Db	632	CGGCGCGCGCATTAAGCGCGCGCTGTGTGGCGCATCAAGAAGAGTACGACACAA	691	
QY	472	CTACAGCTTGATTAAGCGCGCGCTGTGTGGCGCATCAAGAAGAGTACGACACAA	531	
Db	692	GTACACACGACGAAAGGGCGCGCGCTGTGTGGCGCATCAAGAAGAAATACGCAACAG	751	
QY	532	ATCAGCTGGCGAGACCTGATATCTGCGCGCGACCGCTGGTGTAGCTCAGTGGGCTTA	591	
Db	752	ATTCTCTGGCGCGACCTTATCGCGCTCGCGCGACGATGCGCTATGACGTGCGGGCTG	811	
QY	592	CTGCTGTACGCGTTCTTTCTGGCGCGCTCGATATTTGGGAACCCGAAAAGATATCTAC	651	
Db	812	AAGACGCTTGGCTTTCGCGCTTGGCGCGCGCGATATCTGGCGCGCGCGCGCGCGCGCG	871	
QY	652	GTGGGTGAGGAGACCTGCGCGCGCGCGCGCGCTGTGACGCGCGCGCGCGCGCGCGCG	711	
Db	872	TGSGGTGACGAAAGAGTGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	931	
QY	712	CCAGAGACATGGAAGAACCGCGTGGCGCGCGTCCAAATGGGCTGTGATCTATGTGAACCG	771	
Db	932	CCGAGAGCGTGGAAAGACCGCGTGGCGCGCGTCCAAATGGGCTGTGATCTATGTGAACCG	991	
QY	772	GAGCTGTATGACCG	831	
Db	992	GAGGTGTGACGCGCAAGTCCGATCGCGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCG	1051	
QY	832	CGCGCTTATGGCGGATGAACGACGAAAAAACCGACGCGCTTCAAGCTGGCGCGCGCGCG	891	
Db	1052	CGCGCGCGCGGATGAACGACGAAAAAACCGTGGCGCGCGCGCGCGCGCGCGCGCGCG	1111	
QY	892	GATATGTCGCGGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	951	
Db	1112	GCGAGCTCCATGCGCGGATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1171	

QY 952 GTTGAAACACAGCGCTTAGCTGGGCAACCCACACATGAGGGCAAGCAACGACGCC 1011  
 Db 1172 CGGGAATATCAGGGCTCTGGCTGGATCATACCAAGGGCGCGGCTATGGCGTGACAC 1231  
 QY 1012 GTGACTCGGGTATCGAAGGTGCTTTGGGACCAACCAACCCACCAATATCGATGGCTAT 1071  
 Db 1232 GTGTGTGTGGGTTTCGAGCGGCATGTCGACCAAGGCAACCAACAAATGGGACGAGCTTC 1291  
 QY 1072 TTGACGATCGCTGGCTTCACATGGGGATGGAATGGAAGAGTCTCGCGGTGCGAGAT 1131  
 Db 1292 TTGACGATGCTTTCAAGACAGAGTGGACCTTGACGACACGCCCGCGGGTGCATCGCA 1351  
 QY 1132 TGGGAACGATATGACATCAAAAGGAAGAAACAGCGGTGACGCCAGCGACCCCTCTATT 1191  
 Db 1352 TGGGGCGGATATTCACATGATCGGCAAGAGAACAGCTGTGTATGTGCGAGATCGGTGATC 1411  
 QY 1192 GCGCACACCGCATATGACGACGCGGATATGGGATTAAGGTAAATCGACTATCGC 1251  
 Db 1412 CGCACATATCCGATGATGACGACGCGGCACATAGGCTGAGGTGCATCCGATCTACCGC 1471  
 QY 1252 GCTATCTGCGAATAATATGCGCGATCTGTGAGTACTTCMAGAAATATTTGCGAGAGCG 1311  
 Db 1472 GAGATTTGCGTCTGAAGTTCACAGAGATCGGCGGCAAAATGAGTCTGTCTGCTTCGCGCGCGC 1531  
 QY 1312 TGGTGTGACGACGACGCTGAGCTGCGGCGCAAAATGAGTCTGTGATGCTTCGCGCGCGC 1371  
 Db 1532 TGGTTCAGCTGACGATCGGACATGGGCGGCAAGTTCGCTTACCTGCGCGCGGATGTT 1591  
 QY 1372 CCGGACAGACCTGATTTGGCAGACCGGATTCGCGGAGTAAACGACACTCTACGAA 1431  
 Db 1592 CGCGCTGAGACCTGATCTCGGAGATATGATATCCCGGACGCTCCACGACTAGATGTC 1651  
 QY 1432 GAACTGCTCAAGCAGAAATTCACAGATGGCTCGACATTAAGTCAAGTCTGTCGAC 1491  
 Db 1652 GCTGCGCTCAAGGCTTAAGATGCTGCTTCGCGGCTTTCTGTGCGCGCATCTGGTTCAACC 1711  
 QY 1492 GCTTGGGACAGTCCCGCTACTTATCGGGTTTCGCGATATGCGGGCGGTGTACGGGTGCC 1551  
 Db 1712 GCATGGGACAGTCCCGCTACTTTCGTTGCGGACAGGCGGCGGCGCATGGCGG 1771  
 QY 1552 GCAATTTGCTTGGCCACAGACAGTGGTGGGACGACGAGCGCGGCGCTGCGGAA 1611  
 Db 1772 GATATTTGCTTGGCACGCGCAGAGAGATTTGGGAAGGCAATGAGCGCGCGCTCTTCCGC 1831  
 QY 1612 GTGCTGAGGCTTACGACGAGATCTTTCGCGACCGCGGCGCTAGCATCGGAGGTGATC 1671  
 Db 1832 GTGCTTGGTCTTGGACCGATCTGCGCGACAGCGGTGCGACATCGCGGTATGATC 1891  
 QY 1672 GTTCTGGCGGCTAGCTAGCATGCAAGACCGCGGAAGCAGAGGTTCAGATGTGCGC 1731  
 Db 1892 GTTCTGGCTGCGCATATACCGGCTGGACAGCGCGGGAAGCGGCTGGTTTCGATATGCGC 1951  
 QY 1732 GTTCCCTTCTGAAGGCGCTGGGATATGCGACCGCGGCGAGATACGACGACGACTCTTC 1791  
 Db 1952 GTGCTTGTGGCGCTGTGTGGTGTGCTTGGCGGACAGAGATGCGGACGATTT 2011  
 QY 1792 GCACGCTGGAGCGGTGGCGGATGTGCTTGGCACTGGCAGGAAGAAGATGATGTGGT 1851  
 Db 2012 CGCGCGCTTGCACCGCTGGGGATGTGTTTTCGGCAACTGGGTGAAGAGGACTATGTGCTC 2071  
 QY 1852 AACCGGACAGATGCTGTGATGTGGCGACCTGATGGGTTCACCGCGCGCGGAATG 1911  
 Db 2072 AGCGCGGAGAGTCTGCTGCTGCTGCGGAGAGCTCTTGGCTACAGCGCGGAGCTC 2131  
 QY 1912 ACGGTGCTGTGGCGGCTATGGCGGCTACTTGGGCAACCACTATGTGGCACCAACAGCGC 1971  
 Db 2132 ACGGTGCTGATGGCGGCTGCGGCGCTATCGGGCGCAATTCGCGGCGGTGGCGCGCATGCG 2191  
 QY 1972 GTATTCACGAGTTCGAGGCGCATGTGACGACGACGACTTTTGTGACCTGACCGATG 2031  
 Db 2192 GTCTTCACGATAGCCCGCGGCTTACAGACGCACTTCTTCACGAGCTTGACGCAATG 2251

QY 2032 OGGAACAGCTGGAAGCGCGTAGTAGTACCAACCGCTACGAATCCGCGACCGCAAGACGGT 2091  
 Db 2252 GCTCTATCTGGGCTGCGGACCGGCAACATCTCTATGAGATCGGATCGCGACGCGCGC 2311  
 QY 2092 GCGCTGAAGTGGACCGCTCGCGGCTGGATGTGTATTTGGTTCCAATCGCTACTGCGC 2151  
 Db 2312 CGAGCGAGATATTTGGGCAACCCGCGCTCGATCTGTGTGCTGCGCTCAACTCCATCCTCGCC 2371  
 QY 2152 TCTTTACGGAAGTGTACGCCACGAGGATACGCCGAGAGTTCGTGCAGAGACTGCTGTC 2211  
 Db 2372 GCTTATTCGGAGTTTATGCGGACGCGACACGAGGAAATTTCCGCCGCGACTTCTTT 2431  
 QY 2212 GCGCGCTGGAACCAATGATGTAACCGCGACCGCTTTCGA 2249  
 Db 2432 GCGCGCTGGACGAAGGTGATGAACCGCGACCGCTTTCGA 2469

RESULT 7  
 LOCUS 2961086  
 DEFINITION Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 3/6.  
 ACCESSION AP005086 BA000032  
 VERSION AP005086.1 GI:28808836  
 KEYWORDS  
 SOURCE Vibrio parahaemolyticus  
 ORGANISM Vibrio parahaemolyticus  
 REFERENCES  
 1. Nasu, H., Iida, T., Sugahara, T., Yamaichi, Y., Park, K. S., Yokoyama, K., Makino, K., Shingawa, H. and Honda, T.  
 A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:k6 strains  
 2025086 Microbiol. 38 (6): 2156-2161 (2000)  
 2. 2025086  
 10834969

REFERENCE  
 AUTHORS Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tegomori, K., Iijima, Y., Neijima, M., Nakano, M., Yamashita, A., Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shingawa, H., et al.  
 Genomic sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*  
 Lancet 361 (9359), 743-749 (2003)  
 MEDLINE 22508454  
 PUBMED 12620739

REFERENCE  
 AUTHORS 3 (bases 1 to 296100)  
 1. Makino, K., Kurokawa, K., Makino, K., Yokoyama, K., Yasunaga, T., Tegomori, K., Shingawa, H., Hatohi, M. and Iida, T.  
 Direct Submission  
 Submitted (09-APR-2002) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken.gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)  
 genome, Project  
 genome, Project  
 gastroenteritis isolated from a patient presenting with acute

COMMENT  
 Location/Qualifiers  
 1. 296100  
 /organism="Vibrio parahaemolyticus"  
 /mol\_type="genomic DNA"  
 /strain="O3:k6"  
 /db\_xref="taxid:2210633"  
 /db\_xref="taxon:670"  
 /chromosome="2"  
 57..1031  
 /gene="VPA0602"  
 57..1031  
 /gene="VPA0602"  
 note="similar to GB:AAA25445.1 (J04233) percent identity 28.4"  
 /codon\_start=1

FEATURES  
 source  
 1. 296100  
 /organism="Vibrio parahaemolyticus"  
 /mol\_type="genomic DNA"  
 /strain="O3:k6"  
 /db\_xref="taxid:2210633"  
 /db\_xref="taxon:670"  
 /chromosome="2"  
 57..1031  
 /gene="VPA0602"  
 57..1031  
 /gene="VPA0602"  
 note="similar to GB:AAA25445.1 (J04233) percent identity 28.4"  
 /codon\_start=1





Query Match	48.08;	Score 1085;	DB 1;	Length 296100;	
Best Local Similarity	62.24;	0;	Mismatch 660;	Indels 0;	Gaps 0;
Matches 1481;	Conservative				
QY 112	GTTAAATGCTCTGTATCATGACGCTGCTPACACCTCTGCACCGCTACTTCCACAAAGATTGG	171			
DB 194643	GTTAAATGCTCTGTATCATGACGCTGCTPACACCTCTGCACCGCTACTTCCACAAAGATTGG	194702			
QY 172	TGCGCGGAGGATGACCTGCGATATTTGGATCAGCAGATCCGCAATACCGCATG	231			
DB 194703	TGCGCGGAGGATGACCTGCGATATTTGGATCAGCAGATCCGCAATACCGCATG	194762			
QY 232	GATCCGGATTTCACTACCGCTGAGCAAGTACGCACTGGATTTGAGCGCGCTGAGAAA	291			
DB 194763	GATCCGGATTTCACTACCGCTGAGCAAGTACGCACTGGATTTGAGCGCGCTGAGAAA	194822			
QY 292	GATGTCGACGCTTGATGACGATGACGATGACGATGACGATGACGATGACGATGAC	351			
DB 194823	GATGTCGACGCTTGATGACGATGACGATGACGATGACGATGACGATGACGATGAC	194882			
QY 352	GGGGGTTTGATGATCGCTATGCTGCTGGCATCTGGCATCTGGCATCTGGCATCTGG	411			
DB 194883	GGGGGTTTGATGATCGCTATGCTGCTGGCATCTGGCATCTGGCATCTGGCATCTGG	194942			
QY 412	GCTGGGGGGGTGATGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	471			
DB 194943	GCTGGGGGGGTGATGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	195002			
QY 472	GTACACCTTGATTAAGACCGCCCGCTGCTGCTGGCGGATCAAGAGAGATACGGCAAAA	531			
DB 195003	GTACACCTTGATTAAGACCGCCCGCTGCTGCTGGCGGATCAAGAGAGATACGGCAAAA	195062			
QY 532	ATCAGCTGGGCGAGCCTGATGATCTGGCTGGCCCGCTGGCTATGCTGCTGATGGG	591			
DB 195063	ATCAGCTGGGCGAGCCTGATGATCTGGCTGGCCCGCTGGCTATGCTGCTGATGGG	195122			
QY 592	CCTGCTTACGCTGCTCTCTTTGCGCCCGCTGCTGATATTTGGGAAACCGGAAAGATATCTAC	651			
DB 195123	CCTGCTTACGCTGCTCTCTTTGCGCCCGCTGCTGATATTTGGGAAACCGGAAAGATATCTAC	195182			
QY 652	TGGGCTGACGAAAGAGTGGCTGGCACTCTTGACGACGCTACGGCGACGTGACAG	711			
DB 195183	TGGGCTGACGAAAGAGTGGCTGGCACTCTTGACGACGCTACGGCGACGTGACAG	195242			
QY 712	CCAGAGACATGGAACCCGCTGGCGCTGTCCAAATGGGCTGATCTATGACCCG	771			
DB 195243	GAGCTGATTTAGAGACCCCTTCGACGATGATGATGGCTGATCTATGACCCG	195302			
QY 772	GAAAGGTGTAACCGCACCTGATGCTGGTGAAGACCGCACAGAGTACTTGAACCTTC	831			
DB 195303	GAAAGGTGTAACCGCACCTGATGCTGGTGAAGACCGCACAGAGTACTTGAACCTTC	195362			
QY 832	GGCGGTGATGGCATGACGACGAAACCCGACGCTCTCACAGCTGGCGCCACACCGTC	891			
DB 195363	GGCGGTGATGGCATGACGACGAAACCCGACGCTCTCACAGCTGGCGCCACACCGTC	195422			
QY 892	GATTAATGATGCAATGGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG	951			
DB 195423	GATTAATGATGCAATGGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG	195482			
QY 952	GTGAACACCGAGGCTTGAAGTGGGCAACCCGACGCTCTCACAGCTGGCGCCACACCG	1011			
DB 195483	GTGAACACCGAGGCTTGAAGTGGGCAACCCGACGCTCTCACAGCTGGCGCCACACCG	195542			
QY 1012	GTGAACCTGGGTATCGAAGGTGCTGGACACCAACCCGACGAAATTCGATATGGGCTAT	1071			
DB 195543	GTGAACACCGGTATCGAAGGTGCTGGACACCAACCCGACGAAATTCGATATGGGCTAT	195602			
QY 1072	TTGACAGCTGGGTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG	1131			
DB 195603	TTTCTATGCTCTCAATACGATTTGGGACTGACGAGGACCGACCGCCCGCTGGCA	195662			
QY 1132	TGGAAACCGATTGACATCAAAAAGAAAAAGCCGGTGTACGCCAGACCCCTCTATT	1191			
DB 195663	TGGAAACCGATTGACATCAAAAAGAAAAAGCCGGTGTACGCCAGACCCCTCTATT	195722			
QY 1192	GCCACACCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1251			
DB 195723	GCCCTACTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA	195782			
QY 1252	GCTATCTGCGAAAAATTCATGCGCGCTGCTGATGATGATGATGATGATGATGATGATG	1311			
DB 195783	AGATTTTCAGAGCGCTTTTCAACAGAGCCCTGATATTTGATGATGATGATGATGATGATG	195842			
QY 1312	TGGTTCAAGCTGACGACCGCTGACCTGGCCGCGGCAATCAAGCTTACATGCGCGGAGG	1371			
DB 195843	TGGTTCAAGCTGACCTGACCTGATTTGGGACCAAAATGCTGTACCTTAGCGCCAGATGTA	195902			
QY 1372	CCGACGAGAGACTGATTTGGGACAGCCGATTCGCGAGCTAACACGAGCTACTCGGAA	1431			
DB 195903	CCGACGAGAGACTGATTTGGGACAGCCGATTCGCGAGCTAACACGAGCTACTCGGAA	195962			
QY 1432	GAGTGGTGCAGAGAAATTTGACAAATTTGGCTGAGTACTTGTGATGATGATGATGATG	1491			
DB 195963	GATCTTGTGAAAGCAAAATGACGAGCGCGCTGTCTATGCTGATTTGGTGGCCAC	196022			
QY 1492	GCTTGGGACAGCGCCCTACTTATGCGGTTTCCGATATGCGCGCGGTGTCAACGGTGGC	1551			
DB 196023	GCTTGGGACAGCGCCCTACTTATGCTGTGCTGTGACGCGTGGCGGCAATGGCGG	196082			
QY 1552	CGCTTGGCTTGGCCCGCACAGAGAGTGGGCGGCGCACAGCGCGGCGCTGGCGAA	1611			
DB 196083	AGATTTGGTTGGTCCCGCAAAAGATTTGGCAAGGTACGAGCGACGCGCTTAAGTCT	196142			
QY 1612	GTCTGAGGCTTACGAGAGATCTTCGCGACACCGCGGCTTACATCGCGGAGTGATC	1671			
DB 196143	GTCTGAGGCTTACGAGAGATCTTCGCGACACCGCGGCTTACATCGCGGAGTGATC	196202			
QY 1672	GTCTTGGCGGCTTACGCTGATGATGATGATGATGATGATGATGATGATGATGATG	1731			
DB 196203	GTACTTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	196262			
QY 1732	GTCTTGGCGGCTTACGCTGATGATGATGATGATGATGATGATGATGATGATGATG	1791			
DB 196263	GTCTTGGCGGCTTACGCTGATGATGATGATGATGATGATGATGATGATGATGATG	196322			
QY 1792	GCACCGCTTGGAGCGCTGGCCGATGCTGGCTGGCACTGGCAGAGAAAGATGATGGTG	1851			
DB 196323	GAGTGGCTTGAACCTTGGCTGGCGGCTTGGCAAAATTTGGTGAAGAAAGATGATGGTG	196382			
QY 1852	AAAGCGCGAGGCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATG	1911			
DB 196383	AAAGCGCGAGGCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATG	196442			
QY 1912	ACGCTGCTGCTGGCGGATGATGCGGCTACTGGGCGCAACTATGTTGGCAGCAACAGCG	1971			
DB 196443	ACGCTGCTGCTGGCGGATGATGCGGCTACTGGGCGCAACTATGTTGGCAGCAACAGCG	196502			
QY 1972	GATTTACGAGCTTTCAGAGCGGCTGACGACGAGCTTTTGTGAGCTTACGAGCTGAG	2031			
DB 196503	GATTTACGAGCTTTCAGAGCGGCTGACGACGAGCTTTTGTGAGCTTACGAGCTGAG	196562			
QY 2032	GGGACAGCTGGAAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATG	2091			
DB 196563	GGGACAGCTGGAAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATG	196622			
QY 2092	GCGTGAAGTGGGCTGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATG	2151			
DB 196623	CAAGTGAATGAGACCGGCTGATGATGATGATGATGATGATGATGATGATGATGATG	196682			
QY 2152	TCTTACGAGAGAGTGTACGCCCGAGAGATTAACGGCGGAGAGTGTGTCAGAGAGTCTG	2211			
DB 196683	TCTTACGAGAGAGTGTACGCCCGAGAGATTAACGGCGGAGAGTGTGTCAGAGAGTCTG	196742			
QY 2212	GCGCGCTGGACCAAGCTGATGACCGCGGCGGCTTTCGAGCT	2252			

```

DB      196743  GCTGCAATGCAGGAAGTATCATGACGGCGATCGTTTTTGACTTT 196783

RESULT 8
LOCUS   AF197161                                           BCT 14-SEP-2000
DEFINITION Synchococcus PCC6301 catalase-peroxidase gene, complete cds.
ACCESSION AF197161 GI:6180192
KEYWORDS  Synchococcus sp. PCC 6301
SOURCE    Synchococcus sp. PCC 6301
ORGANISM  Synchococcus sp. PCC 6301
REFERENCE Engler,M., Regelsberger,G., Jakopitsch,C., Furtmuller,P.G.,
AUTHORS   Rucker,F., Peschek,G.A. and Obinger,C.
TITLE     Nucleotide sequence analysis, overexpression in Escherichia coli
           and kinetic characterization of Anacyctis nidulans
           catalase-peroxidase
JOURNAL   Biochimie 82 (3), 211-219 (2000)
MEDLINE   20322938
PUBMED   10863004

REFERENCE Engler,M., Regelsberger,G., Jakopitsch C., Furtmueller,P.G.,
AUTHORS   Rucker.F., Peschek.G.A. and Obinger.C.
TITLE     Stimulation
JOURNAL   Substanz U 1999) Institute of Chemistry, University of
           Agricultural Sciences, Muthausse 18, Vienna A-1190, Austria
FEATURES
         source
            1..2700
                /organism="Synchococcus sp. PCC 6301"
                /mol_type="genomic DNA"
                /strain="PCC 6301"
                /db_xref="taxon:1139"
                /date_annotated="2000"
                /note="Anacyctis nidulans"
                /ccds_number="1..11.1.6"
                /note="hydroperoxidase"
                /codon_start=1
                /product="catalase-peroxidase"
                /protein_id="GI_6180193.1"
                /length=2700
                /translation="MTATGRKGVPMHGSGATTNISTADWPKALNDILDSQDRKNP
                MGPFENVEVKLLDVAALKLQDLQALQSTDDMPADMGHVGGLMRLIHTMRAGTYIA
                ADRGAGGTGNORFAPNSPDNTLNOKLRLLRPITQKYGNKSNAWLTAIVAGTYIA
                ESMGLATTCGAFGRDIWIPEKIOTYWGPEKVPVPTKSNYSKRELEPNLAAYTM
                GLTVNPBGADVGGDKPLTKTDGVTGYTFARMANDDEETVALTAGCTVKGCHNGNAL
                LQPPEPGADVGGDKPLTKTDGVTGYTFARMANDDEETVALTAGCTVKGCHNGNAL
                QPPEPGADVGGDKPLTKTDGVTGYTFARMANDDEETVALTAGCTVKGCHNGNAL
                QPPEPGADVGGDKPLTKTDGVTGYTFARMANDDEETVALTAGCTVKGCHNGNAL
                RIASGLISLSLYSWMSDANTSYNSDKGANGARIELARQDMESGEDRLAFLVIA
                VLESIMATGSYADVAVTLACNGVQBOAARGVVELVLPARGDATASQTDDIESPA
                VLEGTHGTGRNMLQDVATPEELLDTOLLIQTAPENTVILIGLVRYGTPNHGTHK
                GVFTDREGVLINTDFEYLMIDMKIPACKNKNIIDR*"
BASE COUNT 702 a 627 c 754 g 617 t
ORIGIN

Query Match          47.68; Score 1075; Db 1; Length 2700;
Best Local Similarity 68.7%; Pred. No. 7.6e-215;
Matches 1481; Conservative 0; Mismatches 675; Indels 0; Gaps 0;

QY 99  GAGACGTATGAGCGGAATGTCCTGTCATGCAGCGGGTGAACAACCTCGACCGGTACTTC TC 158
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243  GACACAGACC CGCGGGTGAATGTCCTGCGGTGCACGACGCGGACGACAAACCACTTAATTTC TC 302
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 159  CACAAAGAATGTGCGCGGAAGGGTTCACACTGCGATATTCCTCAGCACGATCGCAA CTC 418
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 303  GACTCGTGTGTTGGCGCAAGGCACTCAACTCGATGATTTTTCGACCCACGACGATCGCAA CTC 362
       ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 219  ATCAACGCCCGATGGATTCGGATTTTCAACTACGCGTGAAGAAGTACCSAAGTCGATTCGA CTC 278

```





Db 109988 GGCGGTGGTGGCGGATACCGGATTAACCAAGTTTTTGTCTGGTGAATCTTGGCGTGA 110057  
 Qy 469 AAGTGTACGCTGGATTAAGAGCGCGCTGTCTGTGGGCGATCAGAAAGATACGCGAAC 528  
 Db 110058 AACGGACACCTTGATTAAGAGCGAGCTCTACTTTGGCCATTTAGCAAAATAGCGGCAC 110117  
 Qy 529 AAAATCAGCTGGGACGACGTGATGATATCTGGCTGGCAGCGTGGCTATGATGATCAATGGCG 588  
 Db 110118 AAAATTTAGCTGGGCTGATTTGATGATCTTGGCGGTACATGCGGTACGATCAATGGGT 110177  
 Qy 589 TTAGCTGGTATAGCTGTCTGTGGCGCGCTGTGATATTTGGGACCGGAAAGATATC 648  
 Db 110178 CTGAAGAACCTTTGGTTTGGCTTGGCTGTGAAGACATCTGCAATCCAGAAAAGACATT 110237  
 Qy 649 TACTGGGCTGAGAAAAGATGGCTGGACCTTTCTTGCAAAACGCTACGGCAGCTGAAC 708  
 Db 110238 TACTGGGCTGAGAGCAAGATGGTTAGCGGCATCTTGAGGCAAAACAGCAGCTACTCT 110297  
 Qy 709 AAGCAGCAAGCTAGCAAAAGATGGTTAGCGGCATCTTGAGGCAAAACAGCAGCTACTCT 768  
 Db 110298 GTGGAACCGACCTAGAAAACCGCTGCTGTGTGATGATGGCGCTCATTTAGCTGACAC 110357  
 Qy 769 CCGAAGATGTTAAGCGCGACCTGATCGCTGTGAAGACCGCACAGCAGGTACTTGAACCC 828  
 Db 110358 CTGAAGATGTTAGATGTTATTCGCCACCTATAAACAGCAAGATGCTGTGTGACT 110417  
 Qy 829 TTGCGCGTATGAGATGAGCAAGCAAAAGCAAGCGCCCTCACAGCTTGGCGGCCACAC 888  
 Db 110418 TTGCTCTGCATGGGATAGCATATGAAGAAAGCGTTCGCGCTGAGCGCGGTGGTCAACG 110477  
 Qy 889 GTGCGTAATTTGTCACGATAGTGCATGCTCTGTGGTGTAGCCCTTGACCAAAAGCGCTCT 948  
 Db 110478 GTTGTGAAACACAGCTACGCGACCGCAAGCAACTTAGTGGCGGTCCGACAAAGTGGG 110537  
 Qy 949 GAGTGTGAACAGAGCTTAGTGTGGGCAACCCCAATGCAAGCAAGCAAGCAAGCAAC 1008  
 Db 110538 GATTGTGAAGCAAGCGCTAGTGTGAACAAACCAATAGTTCGCGCGGATTTGGTCCGAC 110597  
 Qy 1009 GCCTGTGACTCGGATATGCAAGATGCTGTGGACCAACACCCGACAGAAATTCGATATGGCG 1068  
 Db 110598 AGGTGTACAGTGTATTAAGAGCGCATAGAGCAACCCCAATGAGTAAATCCGACATAT 110657  
 Qy 1069 TATTTCACCTCTGTTCGCTACATATGGGAATGAAAGAGTCTTGGCGGTGCCAC 1128  
 Db 110658 TTCTCTCATCTGATTTAGCTACGATGGTGGTGTACAGAAGCCGACAGCGTCTGGG 110717  
 Qy 1129 CATTTGAGCCCATATGACATCAAAAGCAAAACAGCGGTTTGAACCGCAGCCGCTCT 1188  
 Db 110718 CATTTGAGCCCATATGACATCAAAAGCAAAACAGCGGTTTGAACCGCAGCCGCTCT 110777  
 Qy 1189 ATTTGCGCAACCCGATATGACCGATGCGGATATGCGATTAAGATTAATCCGACATAT 1248  
 Db 110778 ATTTGCGCAACCCGATATGACCGATGCGGATATGCGATTAAGATTAATCCGACATAT 110837  
 Qy 1249 CGCGCTATCTGCGAAATATCATGCGGATCTGTAGTACTTACAGAAACCTTTGGGAAG 1308  
 Db 110838 GGTAAATTTCTGAGCATTCACCAAGATCCGCTTACTTCTGAACTTGTCTGCTCG 110897  
 Qy 1309 CGGTGTTCACATGAGCGACGCTGACCTGGCGGCGGAATACATTTAGCTTACGCGCCGGA 1368  
 Db 110898 GCGTGGTTTAACTGTACTACGCGCATATCGGACAAAGCTGCTACTTGGGCCAGAT 110957  
 Qy 1369 GTTGTGCGCAAGCATTTGATTTGGCAGATCCGATTTGGCGAGGTATACCGACTACTCG 1428  
 Db 110958 GTTGTGCGCAAGCATTTGATTTGGCAGATCCGATTTGGCGAGGTATACCGACTACTCG 111017  
 Qy 1429 GAGAAGTGTCTCAACAGAAATTTGCAAAAGTGGCTGAGCATTTAGTGAATGCTCTCC 1488  
 Db 111018 GTAGATGAGTGAAGCAAAATATCATCGGCAAGTGGTTTGAAGCATTCGCGGAAATGGTTCA 111077  
 Qy 1489 AGCGGTGCGACATGCGGTAGTATTCGCGGTTCGATTCGCGGTGCTGTAGCT 1548  
 Db 111078 ACCGATGGGATAGCGCAGCACTTCCGCAACTCAGATAAACCGGTGGTGGCAAGCGT 111137

RESULT 10  
 AP005085/c  
 LOCUS 303450 bp DNA linear BCT 05-MAR-2003  
 DEFINITION Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 2/6.  
 ACCESSION AP005085 HA000032  
 VERSION AP005085.1 GI:28608465  
 KEYWORDS  
 SOURCE Vibrio parahaemolyticus  
 ORGANISM Vibrio parahaemolyticus  
 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 Vibrionaceae; Vibrio.

REFERENCE 1  
 Masu, H., Iida, T., Sugihara, T., Yamachi, Y., Park, K.S., Yokoyama, K.,  
 Nishio, K., Shingawa, H., and Honda, T.  
 A full genome sequence of Vibrio parahaemolyticus strain BCT 05-MAR-2003 with recent pandemic Vibrio  
 parahaemolyticus O3-K6 strains  
 J. Clin. Microbiol. 38 (6), 2156-2161 (2000)  
 2025086  
 1083469  
 2  
 Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T.,  
 Tsunoda, K., Teraoka, H., Maeno, M., Nakano, M., Yamashita, A.,  
 Kuroda, Y., Kimura, S., Yasunaga, Y., Honda, T., Shingawa, H.,



gene	complement(6548..7201)	QY	652	TGSGGTGACGAAAGAGTGGCTGCGACCTCTCTGACAGCGTACGGGACGTGACGAAG	711
CDS	/gene="VPA0307"				
	complement(6548..7201)	Db	147145	TGSGGTGACGAAAGAGTGGCTGCGACCTCTCTGACAGCGTACGGGACGTGACGAAG	147086
	/gene="similar to GB:A042465.1 (AB002563) percent identity 47 in 200 aa"	QY	712	CGAGAGACCATGGAAACCGCGTGGCGGTCTCCAAATGGGTCTGTGATGTGAGACGG	771
	/codon_start=1	Db	147085	GAACGGGATTTAGAGATCCACTCCCGCTGTGATGTGGCTTAAATTTGTCAACCA	147026
	/evidence=not_experimental				
	/trans_table=11	QY	772	GAGGTGTAAACGGCCACCTGATCCGCTGAGAACCGACGACGAGTACTTGAACCTTC	831
	/product="putative pyrazinamidase/nicotinamidase"	Db	147025	GAGGCGTTCGATGACGACCGGATCACTCAAAACCGCGGACGATGACGATTC	146966
	/protein_id="BAC61650.1"	QY	832	GCCTGCTGCGATGACGACGAAACGACGACGACGACGACGACGACGACGACGACG	891
	/trans_table=11	Db	146965	GCCTGATGCGCAATGAATGATGAGAAACCGTGGCCCTAACGGCTGGTGGCCACACG	147006
	/note="similar to GB:AAF96011.1 (AB004352) percent identity 60 in 241 aa"	QY	892	GTGAATGTGACGGTATGCAATGCGCTAGCCCTGAGCCCTGACCAAAAGCCCTCTGAC	951
	/codon_start=1	Db	146905	GTGAAGCGCTGCTGATGCGTGTGATGACGCTATTTAGCCCGACGACGAGGTGCGAC	146846
	/evidence=not_experimental	QY	952	GTGTAACACCGGCTTAGGTGGCGCACCCACACATCGAGGCGACGACGACGACG	1011
	/trans_table=11	Db	146845	ATTTCACGACCAAGATAGTGTGCTTAAACAGACACCCCGCGCTAGGCAACACGCG	146786
	/product="conserved hypothetical protein"	QY	1012	GTGACCTCGGCTATCGAAGGTGCTTGGCACACCAACCCACGAAATTCGATATGGCTAT	1071
	/protein_id="F28084.1"	Db	146785	GTGACCTCGGCTATCGAAGGTGCTTGGCACACCAACCCACGAAATTCGATATGGCTAT	146726
	/trans_table=11	QY	1072	TTGCGACCTGCTGCTGGCTACATTTGGCAATGAAAGAGTCTCGCGGTGCGCCACAT	1131
	/note="M1VTD1PRLSGCAGKLEVLKSNPRLHYWALPGGF VFDKLTSGGCPADENFEAARIKREKHRYPHFSEAFIDGDKRPDPEDSLMIT HYALVDNNVBOIKNAGVPBQCOLKWFQIQLNGEETLAFQPKTIEMAKOKLASIE YTVSLFDLDEPLVADIISAYQFQIDIRMTIKRLIDSGVLKPTNRVASTNKCG GKPAMVITLSDVTFPQNLG"	Db	146725	TTTCTATCTGCTGTTAAATTTAGATTTGGCAATCTGAAAAAGACCCACGCGCATGGCAA	146666
	45.6% Score 1031 Db 1; Length 303450;	QY	1132	TGSGAACCGATTCACATCAAAAGAAACAGCGGTTGAGCGACGACGCCCTCTATT	1191
	Best local similarity 57.7% Pred. No. 1.5e-205;	Db	146665	TGSGAACCGATTCACATCAAAAGAAACAGCGGTTGAGCGACGACGCCCTCTATT	146606
	Matches 1445; Conservative 0; Mismatches 690; Indels 0; Gaps 0;	QY	1192	GGCCACACCCGATGATGACCGGATGCGGATATGGGATTAAGGTAAATCGACCTATCGC	1251
		Db	146605	GGCCACACCCGATGATGACCGGATGCGGATATGGGATTAAGGTAAATCGACCTATCGC	146546
		QY	1252	GCTACTGCGGCAAAATATCGGCGATCTGATCTACTCAAGAACTTTTCGCGAGCG	1311
		Db	146545	AAATCTCAGAACGCTTCATAGCGACCCGCTTATTCGCGACACCTTTGCTCGGCT	146486
		QY	1312	TGSGTCAAGCTGACGACCGCTGACCTGGGCGCGGAAATACGTTACATCGCGCGGACGTG	1371
		Db	146485	TGSGTCAAGCTGACGACCGCTGACCTGGGCGCGGAAATACGTTACATCGCGCGGACGTG	146426
		QY	1372	CCGCGACAGACCTGATTTGGCAGACCGGATTCGCGAGGTGACGACGATCTAGCGAA	1431
		Db	146425	CCGCGACAGACCTGATTTGGCAGACCGGATTCGCGAGGTGACGACGATCTAGCGAA	146366
		QY	1432	GAGGTGTGACGACGAAATTCGACAAAGTGGCTGAGCATTTAGTCAGATCTCTCAC	1491
		Db	146365	GAGGTGTGACGACGAAATTCGACAAAGTGGCTGAGCATTTAGTCAGATCTCTCAC	146306
		QY	1492	GCTGGGCAAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1551
		Db	146305	GCTGGGCAAGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	146246
		QY	1552	GCGATTCGCTTGGCGCCACAGAGAGAGTGGGCGGACGACGCGGCGGCTTGGCGAA	1611
		Db	146245	GCGATTCGCTTGGCGCCACAGAGAGAGTGGGCGGACGACGCGGCGGCTTGGCGAA	146186
		QY	1612	GTCGTGAGCGCTTACGACGATCTGTCGCCACGACGCGGCTGACGTCGCGAGCGATC	1671
		Db	146185	GTAATACCGCTTTAGAGACATCGCGAAGATGACGCGGCTGCTGCTGCTGCTGCTG	146126
		QY	1672	GTCGTGAGCGCTTACGACGATCTGTCGCCACGACGCGGCTGACGTCGCGAGCGATC	1731
		Db	146125	GTAATACCGCTTTAGAGACATCGCGAAGATGACGCGGCTGCTGCTGCTGCTGCTG	146066

Query Match 45.6% Score 1031 Db 1; Length 303450;

Best local similarity 57.7% Pred. No. 1.5e-205;

Matches 1445; Conservative 0; Mismatches 690; Indels 0; Gaps 0;

QY	112	GTTAAATGCTGTCATGACGGTGTGACACCTGACCGGCTACTTCCACAAAGATGG	171
Db	147685	GGCAATGCGCGGTGAACCTGATATGCTGCTGACGCCATACAGACGAGACGCTGG	147626
QY	172	TGCGCGGAGCGGTGAACCTGATATTTTGGCATGACGACGATGCGCAATGACGCCGATG	231
Db	147625	TGCGCGGAGCGGTGAACCTGATATTTTGGCATGACGACGATGCGCAATGACGCCGATG	147566
QY	232	GATCGGATTTCACTACGCGAGAGAGTACGCGAGCTGATTTTCGACCTGAGMA	291
Db	147565	CGGATGATTTTGAATCCAGAGAGTACCAAAATTCGCTTCTGCACTACCA	147506
QY	292	GATGTCACGCTGATGACGACGATGACGAGGTGGTGGCGGCTGCTGCTGCGGACATAC	351
Db	147505	GACCTCATGCGCATATGACGACGATGACGAGATGAGGCTGCTGCTGCGGACATAT	147446
QY	352	GGCGGTGATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	411
Db	147445	GGCGGTGATGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	147386
QY	412	GCTGGGCGGCGGTGATGACGAGCGAGCGCTTTGCGACCGCTCACTCTGCGCGGACAC	471
Db	147385	GCGCGGCGGCGGTGATGACGAGCGAGCGCTTTGCGACCGCTCACTCTGCGCGGACAC	147326
QY	472	GTGACCTGATGATGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	531
Db	147325	GCTAATCTGACGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	147266
QY	532	ATGACGCTGGCAGACCTGATGATGCTGCTGGCAGCGGCTTACGTCGATGCGCTTAC	591
Db	147265	TTGAGCTGGCGGATCTCACTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	147206
QY	592	CCTGCTACGCGTCTCTTTCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	651
Db	147205	AAAAGCTTGGCTTTCGGATTCGGGCGCGAGACATCTGGCGCCGCGGCTGCTGCTGCTG	147146











gene	/db_xref="GI:24350847"; /translation="MODELINGWHISTIPVSEFMQAPLSTFDELSYSALAPLNRI WDTNLSFGSIVTMTLWFGMGKVIYQQLLADNGSDIWLADARITLAPVTSAPVKVR NLTNLSFGQRCGRKAKYKLYEVLPCDKILCAQDFDGLYVSPKM"; complement(4065..4502) /locus_tag="S04398" /locus_tag="S04398"; /notes="similar to GB:X57352, SP:001628, SP:001629, and PID:23398; identified by sequence similarity; putative"; /codon_start=1 /transl_table=11 /product="conserved hypothetical protein TIGR00256"; /protein_id="AN57365.1"; /db_xref="GI:24350848"; /translation="MIAIDQVSRASVYVNDGTIGADNGLLVLGVGRDNRKMEK LATVNSYVSRDSEKNNLNITQAGSLVWVQITLADYETGLRSPFSGAGTPEQA LGLYEYVAFRCRTGVTGTGQFAADKMYELVNDGPTFHLLQV"; complement(4518..5132) /locus_tag="S04399" complement(4518..5132) /locus_tag="S04399"; /notes="identified by Glimmer2; putative"; /codon_start=1 /transl_table=11 /product="hypothetical protein"; /protein_id="AN57366.1"; /db_xref="GI:24350849"; /translation="MKLAIIAILGITACATPSGAVARGLAKPVSAASTDARA VPAPKVQLAGNTNEVAADALYQALLNLANLYKVPDKVAVQFNNDMLRPSHMA GIDRLNRFYAVHPQJQASPELPAVIGELNKLNFARFVLRQGANVTOGTVTAD KVEILRMLMTDGLAQIAGSLPDPGVPEYLRILPLMPVKS"; complement(5129..6085) /locus_tag="S04400"; complement(5129..6085) /locus_tag="S04400"; /notes="similar to GP:1710151, GB:X62583, and PID:42775; identified by sequence similarity; putative"; /codon_start=1 /transl_table=11 /product="proline iminopeptidase, putative"; /protein_id="AN57367.1"; /db_xref="GI:24350850"; /translation="MLTDVQVAFPIREDTMMGSGQQLHLAQGNPGIPLLYLRGG PGAGASVLEPNFYMLLLDQKAGSLPSGLEHNLHGLICDIEAIRLGI BRCLAGSFTALITYSGELPNRYIAQVIALNPITSEKIDMLYTPSQAQLYPOA QLAQLIEYLVNRYGPNVQADQELNHFERFARWIKELTAGVPIALPRLSGA NHRILIEHISDGLGNSLELSIAVLELQAWWIKRDEL"; complement(6289..7176) /gene="rbn"; /locus_tag="S04401"; complement(6289..7176) /gene="rbn"; /locus_tag="S04401"; /notes="similar to GB:J03161, SP:P11831, and PID:338480; identified by sequence similarity; putative"; /codon_start=1 /transl_table=11 /product="ribonuclease Bn"; /protein_id="AN57368.1"; /db_xref="GI:24350851"; /translation="MTKIKDLAQIRLVLLIGTWIFLHLRQRVDEQDINRACHLAVYT	gene	42.1%; Score 951.2; DB 1; Length 12293; Best Local Similarity 65.7%; Pred. No. 9.5e-189; Matches 1423; Conservative 0; Mismatches 728; Indels 15; Gaps 2;
CDS	11966	11966	GCTTAAGGCTCTTAACCTCGATATCTTGCATCAACAGATCGCAATTCGACCGGTTAG 171
gene	11966	11966	GCTTAAGGCTCTTAACCTCGATATCTTGCATCAACAGATCGCAATTCGACCGGTTAG 171
CDS	11966	11966	GCTTAAGGCTCTTAACCTCGATATCTTGCATCAACAGATCGCAATTCGACCGGTTAG 11907
gene	11966	11966	GCTTAAGGCTCTTAACCTCGATATCTTGCATCAACAGATCGCAATTCGACCGGTTAG 231

Query Match

Best Local similarity 65.7%; Pred. No. 9,5e-189;

Matches 1423; Conservative 0; Mismatches 728; Indels 15; Gaps 2;

112 GGTAAATCTCGTATCGAGCGGTGATACACCTCGACCGGTACTTCCACAAAGATGG 171

11966 GTTAAATGCCCCGTGATCGTGGGCGCATCTCTGGCGTGGCGATACATATGGGACTGG 11907

172 TGGCGGAGGGTGTGAACCTGGATTTTGCATCAGCAGATCCGAAATCAGACCCGATG 231



Qy 351 CGGGGCTTGAATGATTCGATGCTGGCATCTGGCTGGCACTACCGTATTCGTATGG 410  
 Db 689 TGGTGAATATGATTCGATGCTGGCATCTGGCTGGCACTACCGTATTCGTATGG 748  
 Qy 411 CGGTGGGGGGGTGTGATCGGAAGGCGAGCGCTTTGCAACCGCTCAACTCTGGCGGGACAA 470  
 Db 749 TGGGGGTGGGGGGCGCATGCGCAACCAAGCGCTTTGGCCCGCTGAAATAGTGGCCAGATAA 808  
 Qy 471 GCTCAAGCTGGATGAAGACCGCGGTGTGCTGTGGCCGATGATGAAGAGTATGCGCAACAA 530  
 Db 809 GCTCAANTTAGACAAGCGCGCGCTACTCTGGCCGATGATGAAGAGTATGCGCAATAT 868  
 Qy 531 AATCAAGCTGGCAAGCACTGATGATTCGTGGCAACCGCTGGCTATGAGTCCATGGGCTT 590  
 Db 869 ACTTTCCTAGCGAGATTTGATGATTCGTGGCAACCGCTGGCTATGAGTCCATGGGCTT 928  
 Qy 591 ACTTTCCTAGCGAGTCTTTGCGCGCGCTGATATTTGGGAAGCGCAAGAGATCTA 958  
 Db 929 CAAGGTTTACGGTTTTCGGGGTGGTGGGAAGACATTTGGCACCGGGAAGAGCATTA 988  
 Qy 651 CTGGGGGTGCAAAAGAGTGGCTGGCACTCTCTGACGAACCTACCGCGACGTGACAA 710  
 Db 989 CTGGGGGACGAAGAAGAAATGGGCTTCTTCAGAGGACCATCGCTACGGCAGTG 1045  
 Qy 711 GCGAGACATGCAAGAACCGCTTGGCGGCTGCTCCAAATGGTGTATCTATGAGACCG 770  
 Db 1046 CCGGGAAGCTTAGAAATCCCTTGGCGCGCTACAAATGGGATTAATTAAGTTACCTTAC 1105  
 Qy 771 GGAAGGTGTTAAGCGCCACCTGATGCTGGCTGGAGAACCGCACAGCAGGTACTTGAACCTT 830  
 Db 1106 TGAAGGGGTGATGGGACCGGAGTCTTTATGCAAGCCAGGACGTCCGCACTACTCTT 1165  
 Qy 831 CCGCCATGATGGGATGAGCAAGCAAAAGACCGAGCGCTGACAGTGGCGCCACAGCT 890  
 Db 1166 TCCCGCGATGCGCATGAATGACGAGAAACGTTGGCCCTTACCGCTGGCGGGCACACGT 1225  
 Qy 891 CGGTAATGTCAGGTAAAGCAATGCTGCTGGCTAGCCCTGACCCCAAGAAAGCTCTGA 950  
 Db 1226 AGGAAATGCTATGCGCAAGCTTAAGCTGAGCTTATGGGCCGACAGGAGGAGCTGA 1285  
 Qy 951 CGTTGAAACACAGGCTTAGGTTGGGCACTCCCAACATGCAAGGCGAAGCAAGCAAGC 1010  
 Db 1286 TGTGGTGGAGCAGGCTTGGGTTGGCAACCAACCAAGGCAAGGCTGTGGGGGGGAAAC 1345  
 Qy 1011 CGTGACCTCGGTATGCAAGGTGCTTGGACCAACCAACCAAGTTCGATATGGCTA 1070  
 Db 1346 TATGTCAGTGGCATATGAGGGCTTTGGCAACCAACCAAGTTCGATATGGCTA 1405  
 Qy 1071 TATGTCAGTGGCATATGAGGGCTTTGGCAACCAACCAAGTTCGATATGGCTA 1130  
 Db 1406 TTTTATATGTGTTCACCAACCAAGTGGCAATGAAGAAAGTCTTGGCGGGGCTTGGCA 1465  
 Qy 1131 TTGGGAACCGATTGATCAAAAGAGAAACAGCCGGGTGAGCCAGGAGCCCTCTAT 1190  
 Db 1466 ATGGGAACCGGTGATTAAGAGAGATTAAGCCCTGATGTGGAGAGCCAGATAT 1525  
 Qy 1191 TGGCCACACCGATGATGACGAGGAGATATGGCGATTAAGGTAAATTCGACGATATG 1250  
 Db 1526 TGGCCACACCGATGATGACGAGGATATGGCGATTAAGGTAAATTCGACGATATG 1585  
 Qy 1251 GCGTATGCGAAMATTCATGCGCGATCTTGGATCTACTACGAAGAACTTTCGCGAAGCG 1310  
 Db 1586 ACMAATTTTCGAGGATCTTATGCGGAGCGGATTAATTCGCGAAGTATTCGCAAGG 1645  
 Qy 1311 GTGGTCAAGCTGACGACGATGCTGGCGCGGAAATCGAGTACATACGCGCGGAGCT 1370  
 Db 1646 CTGGTTTAACTAATCTACAGGAGTCTAGGACAAAGAGCGCTTACCTTGGCGCGGATGT 1705  
 Qy 1371 GCGGCGACAGAGCTGATTTGGCAAGACCGGATTCGCGAGGTAAACACGACTACTGCA 1430  
 Db 1706 GCGGCGAGAGATTAATTTGGGCAAGACCGGATTCGCGAGGTAAACACGACTACTGCA 1765  
 Qy 1431 AG-----ANGTGGTCAGCAAGAAATTCGCAAAAGTGGCTGAGGCTAGTGAAT 1481

Db 1766 AGGGGAATTAGAGAGCTTGAACAAATTTTAGCTCGGGCTCACCGCTTTCGCAACT 1825  
 Qy 1482 GGTCTCCACGCTTGGGAGAGTCCGCTACTATATCGGGTATCGGATATCGCGGGGCG 1541  
 Db 1825 AGTCTGCTGCTTGGATAGTGGCCGACCTTCGCTAGCTCGAGTATCGAGGGGAGC 1885  
 Qy 1542 TAAAGGTGCTCGATCTCGTTTGGCCGCAACAGAGTGGAGGGCAACAGCGCGGAGCG 1601  
 Db 1886 CATTGAGCGCGATTTAGAGCCCAAAAGATTTGGCGGCTGAATGAACCAACTGAG 1945  
 Qy 1601 CCTGCGAAGCTGTGAGCGCTTAGAGAGAGATCTCTCC-----CGACCCGGGCTG 1655  
 Db 1946 ATTTGGCAAACTTTGGCGGTGCTGGAAATATCCAAAGCTAATTTTCTTAAAGCGGTCAG 2005  
 Qy 1656 CATCGGAGAGTATGCTTCTGGCGGTGAGCTAGGATAGAGAAAGCGGGAAGAGC 1715  
 Db 2006 TCTAGCGGCTTAATTTGCTTAGAGGGGAGCGGCGATTCGCAAGAGGCGGCTTGTATGG 2065  
 Qy 1716 AGSTTACGATGTGGCGGCTTCTTCTGCAAGGCGGTGGGATGCGACGCGAGAGTAC 1775  
 Db 2066 GGGGATCGAGTTAATGTACCTTTCTGCCGCTGCAAGGATGCGCAACCAAGCACTGAC 2125  
 Qy 1776 CGAGCAGACACTTCTTCGCAACGCTGGAGCGGTGGCGGATGAGCTTCGCGCACTGGCAGAA 1835  
 Db 2126 CGATGAGAGATTTTCTTCTTCTGCAAGGCGGTGGGATGCGACGCGAGAGTAC 2185  
 Qy 1836 GAAAGATATGTGTGAGCGCGAGAGAGTCTGATGATGATGATGATGATGATGATGATG 2195  
 Db 2186 ACAGGATATGCTGTGTGCCGCGAGGAATGCTGTGGAGGAGCAACCAATTAATGGGACT 2245  
 Qy 1896 AACCGCGCGGAAATGACCGTCTGCTGGCGGTATGCGGCTACTGGGCAACCACTATGG 1955  
 Db 2246 AACCGCGCGGAAATGACCGTCTGCTGGCGGTATGCGGCTACTGGGCAACCACTATGG 2305  
 Qy 1956 TGGCAGCAACACGCGCTTATCACGATTTGTGAAGCGAGTGTGACCAACGACTTTTGT 2015  
 Db 2306 CGGAGCTAAGCATGAGTGTCTTACCGATGCGGGTGGGAGTACTCATGAGGACTCTTGT 2365  
 Qy 2016 GAACTGACCATTTGGGGAAGCTGGAAGCGGTATGATAGCAACGCTACGAAATCGG 2075  
 Db 2366 TAATTTGACTGATTTGGCTTCCCAATGCGGCGAGCGGCAACATCTCTGAAATGG 2425  
 Qy 2076 CGAGCGCAAGACCGTGGCGTGAAGTGAAGCGCTCGCGGTGATCTGTGATTTGGTTC 2135  
 Db 2426 CGATCGCAACACGAGGAGTGAATGGAATGCCACCAAGTAGACCTGTGTTGGTTC 2485  
 Qy 2136 CAACTGCTACTGCTCTTACGAGAGATGATGCGCCAGAGATACGCGAGAGATTT 2195  
 Db 2486 CAACTGCTACTGCTCTTACGAGAGATGATGCGCCAGAGATACGCGAGAGATTT 2545  
 Qy 2196 CGTCAGAGACTTCTCGCGCGCTGCAACCAAGTATGATGAGCGCGGAGCTTTCAGAGTCG 2255  
 Db 2546 TGTCCGTGATTTTGGCGGCTCGAGCTTAAGTATGATGATGCGGAGCGGCTTGTATCC 2605

RESULT 15  
 D90910/c  
 LOCUS 146174 bp DNA linear BCT 04-JUL-2001  
 DEFINITION Synchocystis sp. PCC 6803 DNA, complete genome, section:12/27,  
 1430419-1576592.  
 ACCESSION D90910 AB001339 BA000022  
 VERSION D90910.1 GI:1652956  
 SOURCE Synchocystis sp. PCC 6803  
 ORGANISM Synchocystis sp. PCC 6803  
 Bacteria; Cyanobacteria; Chroococcales; Synchocystis.  
 1  
 AUTHORS Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Suzuki, T., Miyajima, N.,  
 Sugita, M. and Tabata, S.  
 TITLE Sequence analysis of the genome of the unicellular cyanobacterium  
 Synchocystis sp. strain PCC6803. .1. Sequence features in the 1 Mb  
 region from map positions 64% to 92% of the genome







Db 114965 ATTGCAAAAGTCTTGGCGGCTGGAATAATCAAGCTAATTTTGTAAAGCGGTGAG 114906  
Qy 1656 CATCCCGACGTGACGTCTTCGTCCCGTACGTAGCCATGCGAAGACCGCGCAAGCAGC 1715  
Db 114905 TCTAGCGGACTTAATTTGTCTAGAGGGGAGCGGCATCCCAAGCGCCCTTCATGG 114846  
Qy 1716 AGGTTAGCATGTGCGGCTCCGTTCCCTGAAGGCGGTGGCATGGACCGCGGAGATGAC 1775  
Db 114845 GGGGATCGAGGTTAATGTGACTTTCTCCCGGTGAGGGGATGCCACCAAGCCATGAC 114786  
Qy 1776 CGACCGACACTCTGCGACGCGCGGACCGGCTGCCGATGCGTTCGCGCACTGGCAGAA 1835  
Db 114785 CGATSCAAGTCCGTTACTCCGTTGGACCGATCCACGATGTTATCGTACTGGCTAAA 114726  
Qy 1836 GAAGAGTATGTGTGAAGCGGGAAGAGATGCTGTGATCGTGGCAGCTGATGGGCTT 1895  
Db 114725 ACAGGATACGCTGTGTCCCGGAGGAATGCTGTGGAGGAAACCCAAATTAATGGGACT 114666  
Qy 1896 AACCGCGCGGAANTGACGCTGCTGCGGCGGTATGCGCTACTGGGACGACACTGAGG 1955  
Db 114665 AACGCGACCGGAATGACTCTTTAATCGGTGTATCGGGGTGTGCTACTAACCATGG 114606  
Qy 1956 TGSCACCAACACGCGGTATTACCGGATGTGAAGGCCAGTTGACCAACGACTTTTTTGT 2015  
Db 114605 CGGGACTAAGCATGAGGTCTTTACCGATCGGGTGGGAGTACTCAGTAACGACTCTTTGT 114546  
Qy 2016 GAACCTGACCGCATATGGGGAACGCTCGAAGCGGTAGGTAGCAACGCGCTACGAAATCGG 2075  
Db 114545 TAATTTGACTGATATGGGCTACCAATGCGGCGCCAGCGGCAACAATCTCTACGAATTTGG 114486  
Qy 2076 CGACCGCAAGACGGGTGCGGTGAAGTGGAACGCGCTCGCGGTGGATCTGGTATTTGGTTC 2135  
Db 114485 CGATCGCAACAGGGGAAGTGAATGGACTTCCCAACCAAGTAGACCTGTGTGGTTC 114426  
Qy 2136 CMACTCCGTACTCGGCTTTAGCGAGAGTGTACGCCACGACGATAGCGCGGAGTTC 2195  
Db 114425 CMACTCCATCTCTACGCTCTCTACGCAAGTGTATGCCAGCAGCAATCGGGAAATTT 114366  
Qy 2196 CGTCAGAGACTTCGTGCGCGCTCGACCAAGATGATAGCCCGACGCTTTCGACGTGCG 2255  
Db 114365 TGTCGTGATTTGTGGCGCGCTGGACTAAGTGATGAATGCCGACCGGTTGATCTCCC 114306

Search completed: October 8, 2003, 12:53:40  
Job time : 8212.54 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 05:24:20 ; Search time 581.585 Seconds  
(without alignments)  
10499.111 Million cell updates/sec

Title: US-09-884-889-5  
Perfect score: 2262  
Sequence: 1 atgaataacgcacccgcgtgta.....gtttcgaacgtgcgtgta 2262

Scoring table: IDENTITY\_NWC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :			
M_Geneseq_19Jun03.*			
1:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*		
2:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*		
3:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*		
4:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*		
5:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*		
6:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*		
7:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*		
8:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*		
9:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*		
10:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*		
11:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*		
12:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*		
13:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*		
14:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*		
15:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*		
16:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*		
17:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*		
18:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*		
19:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*		
20:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*		
21:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*		
22:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*		
23:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*		
24:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*		
25:	/SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2262	100.0	2262	19 AAV06554	Alcaligenes (Deley)
2	2262	100.0	2262	24 AB553944	DNA encoding Alca
3	1098	48.5	24 AB078599	Nucleotide sequenc	
4	6703	31.1	2196 11 AAQ06815	Sequence encoding	
5	6723	29.7	24 AA559345	Microscilla furves	
6	6723	29.7	24 AA559345	DNA encoding Micro	
7	620.2	27.4	2331 20 AAQ00818	DNA encoding Micro	
8	618.6	27.3	2331 18 AAT90400	Mycobacterium tube	

9	618.6	27.3	2331	20	AAQ00817	M. tuberculosis ca
C 10	618.6	27.3	423763	24	AAQ051403	Mycobacterium tube
C 11	618.6	27.3	423763	22	AAQ051403	Mycobacterium tube
C 12	618.6	27.3	4411529	22	AAI99682	Mycobacterium tube
13	606.6	26.8	2235	19	AAT89967	Mycobacterium tube
14	578.2	25.6	4795	14	AAQ51531	M. tuberculosis H37
15	578.2	25.6	4795	14	AAQ51531	M. tuberculosis kat
16	578.2	25.6	4795	14	AAQ51531	M. tuberculosis kat
17	412.6	18.2	1852	21	AAQ07490	Fusarium venenatum
18	412.6	18.2	1852	21	AAQ07490	Fusarium venenatum
19	273.2	13.2	532	23	AAQ89654	Agaricus bisporus
20	270.4	12.0	824	23	AAQ73010	DNA encoding novel
21	224.6	9.9	1489	21	AAQ23101	Nucleic acid sequ
22	186.4	8.2	671	21	AAQ23101	Aspergillus oryzae
23	185.2	8.2	660	21	AAQ49864	Mycobacterium tube
C 24	170.2	7.5	6240	23	AAQ50138	DNA encoding novel
25	137.2	6.1	620	17	AAT29130	katG gene fragmen
26	137.2	6.1	620	17	AAT29130	katG gene fragmen
27	135.2	6.0	620	17	AAT29132	katG gene fragmen
28	135.6	6.0	620	17	AAT29131	katG gene fragmen
29	128.6	5.7	4653	23	AAQ73484	DNA encoding novel
C 30	128.6	5.7	4653	23	AAQ89265	DNA encoding novel
C 31	128.6	5.7	4653	23	AAQ89808	DNA encoding novel
C 32	128.6	5.7	4653	23	AAQ90173	DNA encoding novel
33	98.4	4.4	550	24	ABQ37616	Oligonucleotide fo
C 34	98.4	4.4	550	24	ABQ37617	Oligonucleotide fo
C 35	95.8	4.2	150	23	AAQ57617	DNA sequence of ka
36	91	4.0	391	19	AAQ70427	Mycobacterium tube
37	91	4.0	391	24	ABL46037	Mycobacterium tube
38	89.4	4.0	391	19	AAQ70424	DNA sequence of ka
39	89.4	4.0	391	19	AAQ70425	DNA sequence of ka
40	89.4	4.0	391	24	ABL46034	Mycobacterium tube
41	89.4	4.0	391	24	ABL46035	Mycobacterium tube
42	87.8	3.9	331	24	ABL46036	DNA sequence of ka
43	87.8	3.9	331	24	ABL46036	DNA sequence of ka
C 44	85.2	3.8	585	24	ABQ23530	Oligonucleotide fo
C 45	85.2	3.8	585	24	ABQ23531	Oligonucleotide fo

#### ALIGNMENTS

RESULT 1  
AAV06554  
ID AAV06554 standard; DNA: 2262 BP.

AC AC  
AC AAV06554;  
ID 03-JUL-1998 (first entry)

DE DE  
DE Alcaligenes (Deleya), aquamarinus catalase-64CA2 gene.

KW KW  
KW Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;  
pasteurisation; ss.

XX XX  
XX Alcaligenes aquamarinus.

PH PH  
PH Key Location/Qualifiers

FT FT  
FT CDS 1..2262  
/\*tag= a  
/product= "Catalase-64CA2"

XX XX  
XX W09800526-Al.

PD PD  
PD 08-JAN-1998.

XX XX  
XX 03-JUL-1997; 97WO-US16513.

XX XX  
XX 03-JUL-1996; 96US-0574887.

XX XX  
XX (RECO-) RECOMBINANT BIOCATALYSTS INC.

PI PI  
PI Adhikary RS, Robertson DE, Sanyal I;





Db	601	GCTCTCTCTTGGCCGGCTGATATTTGGGAACCCGAAAGATATCTACTGGGGTAC	660	Db	1681	GGTAGCGTAGCGATCGAAGACCGCGAAAGCAGCAGGTTACGATGTGCGGTTCCCTTC	1740
Qy	661	GAAGAAAGAGTGGCTGACGCTTCTGACGACCTGACGAGCTGACGAGCAGCAGAC	720	Qy	1741	CTGAAGAGCGGTGGCGATGTCGACACCGCGAGATGACCGACGACGATCTCTTCGCAACGCTG	1800
Db	661	GCTCTCTCTTGGCCGGCTGATATTTGGGAACCCGAAAGATATCTACTGGGGTAC	720	Db	1741	CTGAAGAGCGGTGGCGATGTCGACACCGCGAGATGACCGACGACGATCTCTTCGCAACGCTG	1800
Qy	721	ATGGAACACCCGCTCGCGGTCTCCAAATGGGCTGTGATATGTGAACCCGGAAGGTGT	780	Qy	1801	GAACCGCTGGCCGATGGCTTCCGCACTGCGCAGAAGAAAGATATGTGTGAACCGGAA	1860
Db	721	ATGGAACACCCGCTCGCGGTCTCCAAATGGGCTGTGATATGTGAACCCGGAAGGTGT	780	Db	1801	GAACCGCTGGCCGATGGCTTCCGCACTGCGCAGAAGAAAGATATGTGTGAACCGGAA	1860
Qy	781	AACGGCCACCTCTGCTGCTGAGAACCGCACAGCAGTACTTGAACCTTCGCCCGTATG	840	Qy	1861	GAGATGCTCTGCTGCTGCTGCGCAGCTGATGGCTTAAACGCGCCGGAATGACCTGTCTG	1920
Db	781	AACGGCCACCTCTGCTGCTGAGAACCGCACAGCAGTACTTGAACCTTCGCCCGTATG	840	Db	1861	GAGATGCTCTGCTGCTGCTGCGCAGCTGATGGCTTAAACGCGCCGGAATGACCTGTCTG	1920
Qy	841	CGCATGACGAGGAAACCCGACGCTCTCAGCTGGCGGCACACCGCTGCGTAAATGT	900	Qy	1921	CTGGCGGCTATGCGCTACTTGGGCAACCACTATGTGTGCGCACAACACGCGGTATTAC	1980
Db	841	CGCATGACGAGGAAACCCGACGCTCTCAGCTGGCGGCACACCGCTGCGTAAATGT	900	Db	1921	CTGGCGGCTATGCGCTACTTGGGCAACCACTATGTGTGCGCACAACACGCGGTATTAC	1980
Qy	901	CACGGTAAATGGAATGCTCTGCTGAGAACCGCACAGCAGTACTTGAACCTTCGCCCGTATG	960	Qy	1981	GATGCTGAAGCGCGATGTCGACGACGACGATGTTTGTGACCTGACCGATATGGGACACAC	2040
Db	901	CACGGTAAATGGAATGCTCTGCTGAGAACCGCACAGCAGTACTTGAACCTTCGCCCGTATG	960	Db	1981	GATGCTGAAGCGCGATGTCGACGACGACGATGTTTGTGACCTGACCGATATGGGACACAC	2040
Qy	961	CAGGCTTAGTGTGGGGAACCCCAACATAGAGGCGAAGCAAGCAGCGCTGACCTCG	1020	Qy	2041	TGGAAGCGGTAGTATGACAGCGCTACGAAATCCGCGACGCCAAGACGGTCCGCTGAAG	2100
Db	961	CAGGCTTAGTGTGGGGAACCCCAACATAGAGGCGAAGCAAGCAGCGCTGACCTCG	1020	Db	2041	TGGAAGCGGTAGTATGACAGCGCTACGAAATCCGCGACGCCAAGACGGTCCGCTGAAG	2100
Qy	1021	GCTATCGAAGGTGCTTGACCAACCAACCGGCTGAGCGCAGACCGCTCTATTCGACCTG	1080	Qy	2101	TGGACCGCTCGCGGTGATGATTTGGTTCACCTGCTACTGGGCTCTTACGCA	2160
Db	1021	GCTATCGAAGGTGCTTGACCAACCAACCGGCTGAGCGCAGACCGCTCTATTCGACCTG	1080	Db	2101	TGGACCGCTCGCGGTGATGATTTGGTTCACCTGCTACTGGGCTCTTACGCA	2160
Qy	1081	CTGCTCGGCTCAATTTGGGAACCTGAAAGAGTCTCGCGGTGCCACACATTTGGGAACCG	1140	Qy	2161	GAAGTGTAGCGCCCAAGCAGATTAACGGCGAGAGTTCGTCAGAGACTTCGTCGCGCGCTGG	2220
Db	1081	CTGCTCGGCTCAATTTGGGAACCTGAAAGAGTCTCGCGGTGCCACACATTTGGGAACCG	1140	Db	2161	GAAGTGTAGCGCCCAAGCAGATTAACGGCGAGAGTTCGTCAGAGACTTCGTCGCGCGCTGG	2220
Qy	1141	ATTGACATCAAAAGAGAGTGTGGACCAACCGGCTGAGCGCAGACCGCTCTATTGCGCACAC	1200	Qy	2221	ACCAAGTGTATGAAGCGCGGACCGCTTTTCGAGCTGCGCTCGTAA	2262
Db	1141	ATTGACATCAAAAGAGAGTGTGGACCAACCGGCTGAGCGCAGACCGCTCTATTGCGCACAC	1200	Db	2221	ACCAAGTGTATGAAGCGCGGACCGCTTTTCGAGCTGCGCTCGTAA	2262
Qy	1201	CGCATGACGAGTGGCGGATGCGGATGAGATTAAGTGAATTCGACCTATGCGCTATCTGC	1260	RESULT 3			
Db	1201	CGCATGACGAGTGGCGGATGCGGATGAGATTAAGTGAATTCGACCTATGCGCTATCTGC	1260	AG78599 standard; DNA, 2839 BP.			
Qy	1261	GAAGAAATCATGCGCGATCTCGTACTCTCAAGAAATCTTCGCAAGCGGTGTTCAAG	1320	AC	ABQ78599;		
Db	1261	GAAGAAATCATGCGCGATCTCGTACTCTCAAGAAATCTTCGCAAGCGGTGTTCAAG	1320	AC	ABQ78599;		
Qy	1321	CTGACGCGCTGACCTGGCGCCGAAATCAAGTACAGTACGCGCGGAGTGGCGGCA	1380	DT	25-NOV-2002 (first entry)		
Db	1321	CTGACGCGCTGACCTGGCGCCGAAATCAAGTACAGTACGCGCGGAGTGGCGGCA	1380	DT	25-NOV-2002 (first entry)		
Qy	1381	GACCTGATTTGGCAAGACCCGATTCGCGCAGGTACACGACTACTTCGGAAGAGTGGTC	1440	XX	Nucleotide sequence of Agrobacterium tumefaciens catalase katA gene.		
Db	1381	GACCTGATTTGGCAAGACCCGATTCGCGCAGGTACACGACTACTTCGGAAGAGTGGTC	1440	XX	Nucleotide sequence of Agrobacterium tumefaciens catalase katA gene.		
Qy	1441	GACCTGATTTGGCAAGACCCGATTCGCGCAGGTACACGACTACTTCGGAAGAGTGGTC	1500	KW	Catalase; KatA; reactive oxygen species; ROS; ROS-inducible promoter;		
Db	1441	GACCTGATTTGGCAAGACCCGATTCGCGCAGGTACACGACTACTTCGGAAGAGTGGTC	1500	KW	Catalase; KatA; reactive oxygen species; ROS; ROS-inducible promoter;		
Qy	1501	AGTGGCGCTGATTTGCGGCTTCCGATATGCGCGGCTGCTAAGCGTCCGCGATTCGC	1560	XX	H202 level; antioxi-dant; gene: as.		
Db	1501	AGTGGCGCTGATTTGCGGCTTCCGATATGCGCGGCTGCTAAGCGTCCGCGATTCGC	1560	XX	H202 level; antioxi-dant; gene: as.		
Qy	1561	TTGGCGCCACAGAGTGGCGGACGACGCGGCGGCTGGCGGAAGTCTGGAC	1620	OS	Agrobacterium tumefaciens.		
Db	1561	TTGGCGCCACAGAGTGGCGGACGACGCGGCGGCTGGCGGAAGTCTGGAC	1620	OS	Agrobacterium tumefaciens.		
Qy	1621	GTCATACGACAGATCTCTCGCAGACCGCGGCTAGCATCGCGAGCTGATCTTCGCGC	1680	PH	Key		
Db	1621	GTCATACGACAGATCTCTCGCAGACCGCGGCTAGCATCGCGAGCTGATCTTCGCGC	1680	PH	Key		
Qy	1681	GCTAGGCTAGGCAATGAGACCGGCGGAGTACGATGCGGCTTCCCTTC	1740	FT	CDS		
Db	1681	GCTAGGCTAGGCAATGAGACCGGCGGAGTACGATGCGGCTTCCCTTC	1740	FT	CDS		

## RESULT 3

AG78599

standard; DNA, 2839 BP.

ABQ78599;

25-NOV-2002 (first entry)

Nucleotide sequence of Agrobacterium tumefaciens catalase katA gene.

Catalase; KatA; reactive oxygen species; ROS; ROS-inducible promoter;

H202 level; antioxi-dant; gene: as.

Agrobacterium tumefaciens.

Key

CDS

Location/Qualifiers

308..2479

/Product= "catalase katA"

WO200263032-A1.

15-AUG-2002.

05-FEB-2002; 2002MO-SG000018.

05-FEB-2001; 2001US-266657P.

(PANS/) PAN S Q.

Pan SQ;

WPI: 2002-643421/69.

P-PSDB; ABB78212.



Db	2072	AGCCCGGAAGAGCTGCTGCTGATGCGGACAGAGCTTCTTGCCCTCACCGCGCGGAACTC	131
Qy	1912	ACGGCTGCTGCTGCGCGCTATGCGCGCTACTGCGGACAGCTATGCGGACCAACAGCGC	191
Db	2132	ACGCTCTCATTCCGCGCGCTGCGCGCTCATCGCGCGCAATACCGCGCGTGGCGCATGGC	2191
Qy	1972	GTATACACGATTTGAGAGCGAGTGTACCAAGCAGATTTTTTGTGAACCTGACCGATATG	2031
Db	2192	GTCTTCACCGATTAAGCCGCGGGCGGCTTACACAGCGATCTTTCAGACAGCTTGAGCGACATG	2251
Qy	2032	GGAGACAGCTGGAAGCCGCTAGTGTAGGACAGCCGTACGAAATTCGCGACCCGACAGCGGT	2091
Db	2252	GCTATTCTCTGGTCCGACCGCGCAACATCTATGAGATCGGTGATGCGAGACGCGC	2311
Qy	2092	CGCGTGAAGTGAGCCGCTCGCGGCTGGATCTGTGTTATTTGTTCCAACTGGCTACTCGCC	2151
Db	2312	GAGCGACGATTTGCGCAACCCGCTCGATCTGTGATCTGGCTCCAACTCCATCTGCGC	2371
Qy	2152	TCTTACAGAGTCTTACGCCAGCAGTACGCGAGGATGAGCGAGCTTCGTCGAGAGCTGCTC	2211
Db	2372	GCTTATGCGGAATTTATGCGACGAGCAGCAACGAGGAAATTTCCGCGGACTTCAT	2431
Qy	2212	CGCGCTCGGACAAAGTGTAGTGAACCCGACGCGCTTTTCCA	2249
Db	2432	CGCGCTGAGACGAGGTGATGATGACCGGACCGCTTTCCA	2469
RESULT 4			
AAQ06815			
ID	AAQ06815	standard; DNA; 2196 BP.	
AC	AAQ06815;		
CC			
DT			
XX	06-MAR-1991	(first entry)	
XX	Sequence encoding heat resistant peroxidase.		
DE	pod10; E.coli UM228; da.		
XX			
XX	Bacillus stearothermophilus.		
Key	Location/Qualifiers		
PH	1..2193		
FT	/*Tag= a		
PN	JP0268684-A.		
XX			
XX	02-NOV-1990.		
XX			
PF	07-APR-1989; 89JP-0089469.		
XX			
PR	07-APR-1989; 89JP-0089469.		
XX			
PA	(TOYM) TOYOBO KK.		
XX			
XX	WPI; 1990-372808/50.		
DR	P-FDB; AA068205.		
XX			
PT	DNA contg. genetic information of heat resistant peroxidase -		
PT	prep. by culturing transformant and collecting heat resistant		
PT	peroxidase		
XX			
XX	disclosure; Fig 2; 8pp; Japanese.		
CC	Sequence may be used to construct plasmid pod10 used to transform		
CC	E.coli UM228. The transformed expression system may be used to		
CC	produce heat resistant peroxidase in large quantities.		
XX			
SQ	Sequence 2196 BP; 576 A; 564 C; 633 G; 423 T; 0 other;		
Query Match			
Query	11; Score 703; DB 11;		
Location	Similarity 61.1%;		
Matches	1326; Conservative 0; Mismatches 765; Indels 44; Gaps 12;		

Qy	138	TAAACGCTCCAGCGTACTTCCACAAAGATTTGGTGGCGGAGGGTGAACCTGGATAT	197
Db	60	TCAGCTCTTCGATGTAAGAGCAGCAAGATCTGGTGGCGGACCACTGAACCTTAAGCAT	119
Qy	198	TTTGATCATCAGCAAGATCGCAATATGAGCCGATGAGTTCGGATTTCAACTACCGTGAACA	257
Db	120	TGTCATCATCATGACGCAAAAACGAATCTCATGATGAGAGTTTCAACTATGCTGAGGA	179
Qy	258	AGTACGCAAGCTCGATTTTCGACCGCTGGAAGAAGATGTCACACGCTTGATGACCGATAG	317
Db	180	GTTCCTCAAACTAGCTATTTGGGCGCTCGAAGAAGATTTGGCAAACTGATGACGGAAG	239
Qy	318	CAGAGATGGTGGCGCTGACTGAGCTGGGCGCTACGSGCGATTTGATGATCGTATGCGCTG	377
Db	240	CCAAAGACTGCTGGGCGCGCGATTAATGGCCATTAACGGCGCTTTGTTATTCGCTGCGCTG	299
Qy	378	GCATCGCTGGCACTACGCTATGCTGATGTCGCGTGGGCGCGTGGTACCGGAGACCA	437
Db	300	GCATTCATTTGCGAGCTACGCTATGCGGACGCGCGCGCGCGCTTCGACGGGACCA	359
Qy	438	GGCTTTGACACCGCTCACTCTCGCGGACACGTCAGCTGGATTAAGCGCGCGCTCT	497
Db	360	GGCTTTGCGCGCTTAACAGCTTCGCGGACACGCCACTTGGATTAAGCGCGCT-T	418
Qy	418	GCTGTGGCGGATCAAGAGAGTACGCGAACAAATCAGCTGGCGAGCTGATGATCT	557
Db	419	GTATGAGCGCATCAAAAGAAATACGGGACAAATCTCTTGGCGCGATTTGCTATTT	478
Qy	558	GCTGCGACCGCTGGCTTATGAGTCCATGGGCTTACTGCTTACGGCTTCTTTGGCGCG	617
Db	479	GGCGCGCATGTGCTGCTATTGATCGATGGGTGGAAAAACGATTGGAATGGCGCGCGCG	538
Qy	618	GTCGATATTGGGACCCGAAAGATATCTACTGGGTGACGAGAAAGATGGCTGGCG	677
Db	539	GCTTGAAGCTCGCATCGGACGAGACGTTATGGGATCGGAAAGATGGCTGGCTGGC	598
Qy	678	ACCTTCTGACGAAAGCTTACGGCGACGTAACACGACGACACATGAAACACCGCTGGC	737
Db	599	CTCT-----GAAGCTATTCCGGGTATGCGGA-----CTCGAAACCGCTCGC	643
Qy	738	GCTGTGCAATTTGGGTCTGATCTATGTAACCGGAGATGTTTACGGCGCACCTGATCC	797
Db	644	CGC-GTGCATTTGGGTATCTCTACGTCAACCCAGAGAGCGCGCAGCGGATGCC	702
Qy	798	GCTGAGAACCGCACAGAGCTACTTGAACCTTCGCGCGCTATGCGGATGACGACGAAAA	857
Db	703	AAAGCGAGGGGATATCG-----CAGAGACGTTCCGCGCGATGGGATGAGGATGAGA	758
Qy	858	TAAGCGAGCGCTCAGCTGCGGCGGACGCGTGGATGTCACGGTATGCGGATGCGCATC	917
Db	759	AAAGCTCGCTTGATCGCGCGGTCTATACGTTCCGAAAGACGACCGCGCGCGCTCGC	818
Qy	918	CTCTGCTTAGCCCTGACCCAAAGACCTCTGACGTTGAAAAACCGAGGCTTAGGTGGGG	977
Db	819	CAGCCATCGCTGCTCCGCGGAGAGCGCGCGCATGAAAGCGCAAGGGCTGGGATGAG	878
Qy	978	CAAGCCACATGCAAGCGGACGACGCGCTGACCTCGGCTATCGGCTATCGGATGCGATG	1037
Db	879	CAGCTTTAGCGAAAGGAGGAGCGGATGCAATGATGATGATGCGCTGACGAGGCGTGG	938
Qy	1038	GACCAACACCCACGAAATTCGATATGCGCTTATTCGACCTCTTCGGCTTCAATATG	1097
Db	939	GAGCTGAAAAGATGCTTCGCGGTGCGCCACCATTCGGAACGATTCGATCAAAAGCA	1157
Qy	1098	GAGCTGAAAAGATGCTTCGCGGTGCGCCACCATTCGGAACGATTCGATCAAAAGCA	1157
Db	999	GTGCGTGAACAAAGACCCACCGCGGCTATGCAATGATGATGCGCTGACGCGGATGAAA	1058
Qy	1158	AAACAGACCGGTTCGACCGCAGCGACCCCTCTATTCGCGCAACACCGCATGACCGATGC	1217
Db	1059	AGACTTGGCTGGATGCGCGAGGATCCGTCGAAAAAAGTTTCCGACGATGATGATGACGAC	1118



QY	1218	GGATATGGCCGATAAGGTAATCCGACCTATCCGCGTATCTCGGAAAATTCATGCGCGA	1277
Db	1119	CGATTTGGCTCGGTGTGACCGCGGANTGCGAAAATTCGCCGCGTTCATCATCAA	1178
QY	1278	TCCTGAGTACTTCAGAAAGATCTTCGGAGAGCGGTGTCAAGCTGAGCGACCGTGACCT	1337
Db	1179	TCGAGAAATTCGCGAGCGGTTCGCGCGGATGTGTCAAGCTTACCATAGGGGAT	1238
QY	1338	GGCGCGAATCATCTATCGCGCGAGATGCGCGGAGAGACCTGATTTGGCAGA	1397
Db	1239	GGGACCGAAGAGATATCTCGCGCGGAGTTCGGGAAGAGATTCATCTGCGCA	1298
QY	1398	CCCGATTCGGGACAGTAACCGCACTACTCGGAA-----GAAGTGTCAACAGAA	1448
Db	1299	TCGATTCGAGATCGATACAGAAATGACAGAGAGGAAATGAGAGAAATCAAGCCAA	1358
QY	1449	ATTTGCGAAMGTGCGTGAAGATGCGTGCATGCGTGTGCGAGCTGCGCG	1508
Db	1359	ATTTTGAATCGGCGCTGACCTCAGTGAATCTGTGAAACAGCTGGG-CCTCGGAG	1417
QY	1509	TACTTATTCGCGGTTCGATATGCGGCGGTGTCAACGGTGGCGGATTCGCTTGGCGCC	1568
Db	1418	CACCTTTCGCGACTCGGATGAGAGGGGG--ACGAAACGCGCGCGATTCGTTCGGCG	1475
QY	1569	ACGAGACATGCGCGGACGACGCGGAGCGCGGCGCGCGGAGTGTGAGCGTCT----	1624
Db	1476	GCAAAAGACTGSGAAGTGAACGACCGAGCGGCGTCCGCAAGTGTGCGCTCTTNG	1535
QY	1625	---ACGACAGATCTCTCGCACCGCGGTGTAGCATCGGAGGTGATGTTCTGGCG	1681
Db	1536	AGGACATCGCGGCGAATCGCCGCAAAAAGTAAGCATCGCGACTGTATGCTTGGCG	1595
QY	1682	GTAGCTAGCATGAGAGAGCGGAGAAACAGACAGTACAGTATGCGGTCTCTTC	1741
Db	1596	CAGCGCTCGG--TGAAAAAGCGACGCGGACGCGCTTGTGATGCAATTCGCATTTT	1653
QY	1742	TGAAGCGCGTGGGATCGACGCGCGGAGATGACGACGAGACTCTTCGACCGCTG	1801
Db	1654	TCCCTGGCGCGGCGATGCGACAGCAAGCAACCGATGTCGNAAGCTTGGCGTGG	1713
QY	1802	AGCGCTGCGCGATGCGCTTCGCACTGCGACGAGAGAGATGTGTGAGCGCGGAG	1861
Db	1714	AACCGTTCGAGTGGCTTCGCAACTATCAAGCAGAGTACACGGTTCGCGCGAAG	1773
QY	1862	AGATGCTGCTGATGTCGCGCATGTATGGCTTACCGCGCGGAAATGACCGTGTGC	1921
Db	1774	TGCGCTGTGTCAGCAAGCCGACTCTCGCG-TGACGCGCGGAGAAATGACGGTCTT	1832
QY	1922	AGCGCGGTGCTGACTGCGCACCATGTGTGGCACCAACAGCGGTATACCG	1981
Db	1833	TTGCGGTTTGGCGGTGTGTGGG--CGAATATCGGCATCTGGCTCATGGGCTTCACT	1891
QY	1982	ATTCGAGGCGAGTGCACCAAGACTTTTGTGTGACCTGACCGCATTTGGGACAGCT	2041
Db	1892	ACCGCATCGGGTGTGACCAAGCACTCTTGTGTCACTTGTGTGATATGAATGAAT	1951
QY	2042	GGAGCGGTAGTAGACACGCGTACGAATCGCGACCGACGACGCGCGCGGAGT	2101
Db	1952	GGGTGCGACAGACAGCGCATTTATGAAATCGCGACCGGAAACGGGCGAGTGGT	2011
QY	2102	GGACCGCTTCGCGGTGGATGTGGTATTTGTGTCCAATCGCTACTGCGCTTACGAG	2161
Db	2012	GGAGCGGACCGGGTGGATCTCATTTTGGATTTTACGCCCAAGCAACCAAGAG	2071
QY	2162	AGTGTGACCGACGACGATAGCGGAGAGTCTCATGAGTGTGCTGGCGCGTGA	2221
Db	2072	TAATTTTACGCCACAGACACCAAGAAATTTGTCCTGTATTCATCAAGCTTGG	2131
QY	2222	CCAAATGATGAACGCGGACCGTTCGAGCTGCGG	2256
Db	2132	TGAAATGTCATGAACCGCGCTTTGATGTGGTG	2166
<hr/>			
RESULT 5			
AAV06555	AAV06555 standard; DNA; 2238 BP.		
AC	AAV06555;		
DT	03-JUL-1998 (first entry)		
DE	Microscilla furvescens catalase-53cAl gene.		
KE	Catalase: epoxidation; hydroxylation; biosensor; paper bleaching;		
KW	pasteurisation; ss.		
OS	Microscilla furvescens.		
XX	Key		
XX	CUS		
XX	Location/Qualifiers		
XX	1..2238		
XX	Product		
XX	/product="catalase-53cAl"		
XX	W09080526-A1.		
XX	08-JAN-1998.		
XX	03-JUL-1997; 97WO-0516513.		
XX	03-JUL-1996; 96US-0674887.		
XX	(RECO-) RECOMBINANT BIOCATALYSIS INC.		
XX	Adhikary RS, Robertson DB, Sanyal I;		
XX	WPI: 1998-086953/08.		
XX	F-F5DB; AAW3810.		
XX	New bacterial catalases, related nucleic acid vectors and		
XX	transformed cells - used as oxidising agents and for detecting or		
XX	destroying hydrogen peroxide, e.g. in biosensors		
XX	Claim 3; Fig 2; 35pp; English.		
XX	The present sequence is of the Microscilla furvescens catalase-53cAl		
XX	gene. Fragments of the gene can be used to identify related sequences.		
XX	Catalase-53cAl may be used to catalyse oxidation reactions such as		
XX	epoxidation or hydroxylation. The enzyme can also be used to detect or		
XX	destroy hydrogen peroxide, e.g. in connection with glyoxylic acid		
XX	production. Biosensors, contact lens cleaning, pulp/paper bleaching and		
XX	pasteurisation of dairy products. Antibodies raised against		
XX	catalase-53cAl can be used to screen libraries for detection and		
XX	purification of cells containing the enzyme.		
XX	Sequence 2238 BP; 634 A; 545 G; 605 G; 454 T; 0 other;		
QY	Query Match		
QY	Beat Local Similarity 59.5%; Score 672.2; DB 19; Length 2238;		
QY	Matches 127; Conservative 0; Mismatches 808; Indels 60; Gaps 6;		
QY	150	CGGTACTTCCAAAGAGTGTGGCGGAGCGGTGACCTGACCTGATATTTTCACCA	209
Db	99	CGGCAACAAAAGGGATTTGGTGGCAACATGCTCAACTCGGCACTTACGCCACA	158
QY	210	AGATCGCAATACAGACCGGATGATCGGATTTCAACTACCTGAGACGATGACCA	269
Db	159	TTCATCGCTATCGGACCAACGACCGGATTTTGACTATGCGGAGAGTTAAGAGCT	218
QY	270	CGATTCGCGCGGTGAGCAAGATGTCCACCGCTTGACGACGACGACGACGACG	329
Db	219	AGATCTGCGACGGTTAAAGAGCACTGCGACGCGCTAATGACAGATTCACAGACT	278
QY	330	GCCCGGTGACTGGGGCACTACGGGGGTTTGATGATCGGTATGCGTGTGGCACT	389
Db	279	GCGACGAGATACGGTCAATTTGGCCCTTCATTTACGATGCGGTGGGACAGCGCG	338

Db	1404	GGATCTGGGACCTGACGCTGAGCGAGCTTGATGACGACGCTTGGCTTGCATCTACTTT	1463
QY	1515	TGCGGGTTCCGAGTATATGCGCGGGTGCTATGCGCGCGGCAATGCGTTGGCCCGACGAA	1574
Db	1464	TGAAATCTCTGACAGACGCGCGGGTGCCACAGCGTGCACGTATAGCATGTGGCCCGACAAA	1523
QY	1575	CGATGTGCAGGGCGACGACGCGGAGCGCTGTGGCGAAAGTGGCTG	1619
Db	1524	AGCTTGGGAGGTATACAAACCTCTACCACTTTTCCGAGGTACTTCAAACTATAGAGGTAT	1583
QY	1620	-----CGTCTACGAGCAGATCTGTGGCGACACAGCGGCTAGCATGGCGACCTGAT	1670
Db	1584	CCAGGAGGACTTTAAACGAGCGCAATCAGATACAAAGACGATATGGTTGGCGACCTGAT	1643
QY	1671	GGTTTGTGCGCGGTAGCTTAGGCAATCGAGAAAGCCCGCAAGACAGCTTTACGATGTGCG	1730
Db	1644	GTGTTTGGCCGCGTGTGCGGGTGTAGAAAGCTCTGCAAGTCTGGATGCTGGCGATGACGA	1703
QY	1731	GTGTTCCCTTCTGAAAGGCGGTGGCGAGTGCACGCCGCGAGATGACGACGACGACCTCT	1790
Db	1704	GGTGGCTTTTCAACCGCGGACGAGCGGATGCCACGCTGAGCAACCGATGTGGAAGCTT	1763
QY	1791	GCACCGCTGAGCCCTGTGCGGATGGCTTCCGCACTGGCAGCAAGAGTATGTGGT	1850
Db	1764	CGAAGCATAGACCGAGCGGTGACGGCTTTAGAACTACATTAACCGGACCTAAAGCT	1823
QY	1851	GAGCCGGAAGATAGTCTGGATCGTGTGCGCAGCTGATGGGGTTAACCGCGCCGGAAT	1910
Db	1824	ATCCGCTGTAGGAATATCTCGTATAGACGGCGGCAAGTCTGTGTGCGCTTTCGGCACCAAT	1883
QY	1911	ACTGGCTGTGCTTGGGGGTGTATGGGCTATGTGGGCAACATCTATGTGGCCACMACGG	1970
Db	1884	GACTCTTTGGTATGCGGTATGCGTGTACTGGCCACCATACGACGGTTTCGACGCTCG	1943
QY	1971	GGATATCACCGATGTGGAAGCGCATGTGACCAACGACTTTTTGTGAACCTGACCGATAT	2030
Db	1944	ATGTTTATCAAAATAGCGGGGTAGCTATCGATGACTTTGTGTAACCTGCTAGACCT	2003
QY	2031	GGGCGACAGCTCG-----AAGCGGTAGTGTAGCAAGCGCTACGAAATGCGGACCGCA	2084
Db	2004	CNACATTAATGCGGACGACGAGTGAATCAGACAAAGTTTTTGAAGCGACGACTTCMA	2063
QY	2085	GACCGGTGCGGTGAAGTGGACGCGCTCGCGGGTGGATCTGGTATTGTTTCCATCTGCT	2144
Db	2064	ACTGGCGAAGTAACGTGGATGGTACCGCGGAGGACCTGATCTTCGAGTTCGATTCGCA	2123
QY	2145	ACTGGCGCTCTAAGCGATGATGTACCGCGGAGTATACGCGAGAGTCTGTGCGAGG	2204
Db	2124	CTAGTAGCCCTCTCGAAGAGTGTAGCGCTGTGCAGATTCTCAAGAAAGTTGTTAAGA	2183
QY	2205	CTTCTGTGCGCGCTGGACCAAAAGTATGATGACGCCGACCGCTTTCGA	2249
Db	2184	TTTTGTGAGGCGCTTGGGCAAGTAATGTGACCTTGGACCGGGTTTGA	2228
RESULT 6			
Db	ABSS3945	ABSS3945 standard; DNA; 2238 bp.	
AC	XX	ABSS3945;	
XX	XX	22-NOV-2002 (first entry)	
DE	XX	DNA encoding Microscilla catalase protein 53kDa.	
XX	XX	Gene; ss; marine bacteria; catalase; hydrogen peroxide; enzyme.	
OS	XX	Microscilla furcascens.	
XX	XX	Location/Qualifiers	
FT	CD5	K89	
FT	FT	21238	
FT	FT	7189= 8	

FT XX /product= "Catalase protein"  
 PN US2002102680-A1.  
 PD 01-AUG-2002.  
 XX 19-JUN-2001; 2001US-0884889.  
 PF 03-JUL-1996; 96US-0674887.  
 PR 16-OCT-1997; 97US-0951844.  
 PR 05-OCT-1999; 99US-0412347.  
 XX (ROBE/) ROBERTSON D E.  
 PA (SANTY/) SANTAL I.  
 PA (ADHI/) ADHIKARI R.  
 PR ROBERTSON DE, Sanjal I, Adhikari R;  
 XX P-PSDB: ABG33041.  
 DR WPI: 2002-690613/74.  
 XX P-PSDB: ABG33041.  
 PR New purified catalase polypeptide useful for catalyzing the breakdown  
 PT of hydrogen peroxide, for modifying small molecules, and for generating  
 PT antibodies which bind to the polypeptide  
 PT Claim 1: Fig 6; 44pp; English.  
 XX This invention relates to the DNA and protein sequences of a novel  
 CC purified catalase polypeptide. The protein of the invention or its  
 CC homologue is useful for catalysing the breakdown of hydrogen peroxide.  
 CC The protein sequence of the invention is useful for modifying small  
 CC molecules, by mixing the protein or its fragments with a small molecule  
 CC and then exposing the mixture to a catalase enzyme. The protein of the invention is  
 CC also useful for catalyzing biochemical reactions, for catalyzing  
 CC glycosidic linkages and for generating antibodies which bind  
 CC specifically to the protein. The nucleic acid sequences of the  
 CC invention is useful as a probe to determine whether a biological sample,  
 CC such as a soil sample, contains an organism having the nucleic acid or  
 CC an organism from which the DNA was obtained, or for identifying an  
 CC organism having the nucleic acid. An antibody which binds the cellulase  
 CC protein of the invention is useful for identifying the presence of the  
 CC protein in a sample. The protein of the invention is also useful for  
 CC procedures to isolate or purify the protein or for detection of protein  
 CC expression in a biological sample. The cellulase protein of the  
 CC invention is heat stable, is heat resistant, and is able to  
 CC renature and regain activity after exposure to temperatures of from  
 CC about 60 to 105 degrees. The present sequence represents the DNA  
 CC sequence encoding the Microcella furvensis catalase protein of the  
 CC invention.  
 XX  
 SQ Sequence 2238 BP; 634 A; 545 C; 605 G; 454 T; 0 other;

Db 339 CACCTACCGTATCGCGTGTGATGGCGTGGCTCGCGCTCAGAGCGCTTGGCGC 398  
 Qy 450 GCTCAACTCTCGCCGCGACAGCTCAGCTCGATTAAGAGCGCGCTGTGCTGCGCGAT 509  
 Db 399 TCTCAATAGCTGCGCAGACATGCCATCTGATTAAGCAGCACCTTGTCTTGGCCAT 458  
 Qy 510 CAAGAAGAGTACGGCAGCAAAATCAGCTGGGCGAGACCTGATGATCTTGGCTGGCACCCT 569  
 Db 459 CAACAAAATACGGTCGGAATAATCTCTGGGCGGATCTAAATGATACTACAGAAACCT 518  
 Qy 570 GCGTATAGCTCCATGGCTTACCTTGTTCAGGCTTCTCTTTCGGCGCGGTGAGATTGG 639  
 Db 519 AGCTCTGGAACATATGGGCTTAAACCTTTTGGTATTGACAGGTGCGAGAGATGATG 578  
 Qy 630 GGAAACCCCAAAAGATATCTACTGGGGTGACGAAAAAGAGTGGCTGGCACTTCTTGACGA 689  
 Db 579 GAGAGCTCAAGAGATGATATCTTGGGAGCAGCAACCGAATGGCTGGG-----GACAA 632  
 Qy 690 AGCTACGGCGACCTGACAGCGACGAGACCATGAGAAACCGCGCGCTGCTCCAAAT 749  
 Db 633 CCCTTATGAAGTGCACCGAGAGC-----TCGAAAATCCCTCGGAGCGCTACAAAT 683  
 Qy 750 GGGTCTGATCTATGTGAACCGGGAAGGTGTTAAGCGGCCACCTCTGATCGCTGAGAGACCG 809  
 Db 684 GGGACTCTATCTATTAACCGCGAAGGACCCAGCGCAACGACCCCTATCTCGTGTGTC 743  
 Qy 810 ACAGCAGGTACTTGAACACTTTCGCCCTATGGGATGAGACGACGAAAAAGCGCGCT 869  
 Db 744 CCGTGATATTCTGAGACTTTTGGCCGATGGCAATGAATGACGAGAAACCTGGCTCT 803  
 Qy 870 CACAGCTGGGCGCACACCGTCTGGTAAATGTTCAGGTAATGGCAATGCTCTCGCT---T 926  
 Db 804 CATAGCGGTGGACACACCTCTGGGAAACCGATGCTGCGGATGCGGAGAAATATGT 863  
 Qy 927 AGCCCTCGACCAAAAGCCTCTGACGCTTGAACACAGCGGCTTNGSTGGGCGACGCCAA 986  
 Db 864 GSGCGAGAGCCTCGCGCGCGAGTATTTATGAGAATAGAGCTGGGGTGGAAAAACACTA 923  
 Qy 987 CATGCGAGGCAAGCAACCGCTGACCTCGGATATGCAAGGTGCTTGGACACCAAA 1046  
 Db 924 GGGACCGCGACAGCTGGCGTACCTACACGTGGCGTGAAGGCGCGCTGGACCAAGAC 983  
 Qy 1047 CCCCACGAATATCGATATGSGCTATTTCCGACTGCTGTTCCGCTACATTTGGGACTGAA 1106  
 Db 984 CCCTACTCAATGGAGACATACACTTTTGTGAAAACCTCTTTGGTACGATGGGAGCTTAC 1043  
 Qy 1107 AAGAGTGTCCGCTGGCCACCTCGGACATCTGGGACAGTTCACATCAAAAAGGAAAAACGCG 1166  
 Db 1044 CAAAGCTCCAGCTGGAGCTTATCTGCTGTAACCAAGAGAGCGTCCGGGCTGGCGACT 1103  
 Qy 1167 GGTTCAGCCACGACCCCTCTATTTCGCCACACCGCATGATCGATCGGATATGCG 1226  
 Db 1104 ACCGATGACATGATGCTCCAGCAAGTCGCGAGCTCCATTATTGCTACTACGAGCCTGCG 1163  
 Qy 1227 GATTAAGGTATTCGACATTCAGCGCTATCGCGCTATCGGAAAAATTCATGGCGGATCTGATGA 1286  
 Db 1287 CTTTCAAGAAAACTTTGGGAAGCGCTGGTTTCAAGCTGACCCACCGTACCTGGGCGCGAA 1346  
 Qy 1224 GTTTTCAGATGTTTGGGAAGCATGTTACAACTGACACACAGATATGGGACCAA 1283  
 Db 1347 ATCACTTTACCTGGCGAGTGGCGGAGAGACCTGATTGGGAAGACCGGATTC 1406  
 Qy 1284 GGTGGCTACTGGGACCAAGATCTCTAGCAGAGAGCTCATCTTGGAGAGACCTTATAC 1343  
 Db 1407 GCGACGTAAC-----ACCGACTACTCGGAGAAAGTGTTCAGAGCCCTAAAGGCCAAATATTC 1454  
 Qy 1344 AGATGTAAAGCATCTCTTGTAGAGCAAAACGATATTCAAGCGCTTAAGGCCAAATCTCT 1403  
 Qy 1455 ACAAGTGGCTGAGCATATGATGAGATGTCTTACCGCTTGGGACAGTGGCGCTACTTA 1514  
 Db 1404 GGGATCGGAGCTGACGCTAAGCGAGCTGTGAAGCGCGCATGGGCTTCTGATCTACTTT 1463





FT	mutation	859	b	*Tag- /note-	*G to C transversion at base 1013 alters codon 315 from AGC (Ser) to ACC (Thr) leading to Isoniazid resistance, and eliminates a CfoI restriction site*	232	GCTGACCCGATGGGTGGCGGCTTCGACTATGCGCCGGAGGTGGGACCATCGACGTTGAC	291
FT						280	GGCTGTGAGAAAGATGTCACGGGTGTGATGACATAGCAAGATGTGGCCCGCTGAC	339
FT						292	GGCTGTGAGAAAGATGTCACGGGTGTGATGACATAGCAAGATGTGGCCCGCTGAC	351
FT	mutation	1013	c	*Tag- /note-	*G to C transversion at base 1013 alters codon 315 from AGC (Ser) to ACC (Thr) leading to Isoniazid resistance, and produces a new MspI restriction site*	340	TGGGGGACTACGGCGGTTTGTGATGATGCGTGTGGCAGCTGGCTGGCGACCTACCGT	399
FT						352	TACGGCCACTACGGCGGCTGTTTATCGGATGGCTGGGACGCTGGCGACCTACCGC	411
FT						400	ATTGCTGATGGCCGTGGGGCGGTGTACCGAAGCCAGCGGTGTGGACCGCTCAACCTCC	459
FT	mutation	1079	d	*Tag- /note-	*A to G transversion at base 1079 alters codon 337 from TAC (Tyr) to TGC (Cys) leading to Isoniazid resistance, and eliminates a RsaI restriction site*	412	ATCCAGACACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACGCG	471
FT						460	TGGCGCGACAAAGTACAGCTGAAGAGCGCGCTGTCTGTGGCGCGCTGATATTGGGAACG	519
FT						472	TGGCGCGACAAAGTACAGCTGAAGAGCGCGCGCGCTGTCTGTGGCGCGCTGATATTGGGAACG	531
FT	mutation	1457	e	*Tag- /note-	*G to T transversion at base 1457 alters codon 463 from CCG (Arg) to CTG (Leu) leading to Isoniazid resistance, and eliminates a NciI-MspI restriction site*	532	TACGGCCAGAACTGCATGGCGGACCTGATTTTCCCGCGACACTGCGCGCTTACGCG	591
FT						580	TGCATGGGCTTACTGCTTACGGCTCTCTTTGGCGCGCTGATATTGGGAACG	639
FT						592	TGCATGGGCTTACAGAGCTTCGGGCTTGGCGCGGTGACACAGTGGGAGCGCG--	649
PN	US558733-A.					640	AAAGATATCTTCTGGGTGACGAAAGAGTGGCGGCACTTCTGAGGAAAGCTTACGCG	699
XP	19-AUG-1997.					650	ATGAGGTCTATTGGGCGAAGGACCACCTGGCTCG-----GCGATGAGGTTCACGC	702
XX	07-APR-1995; 95US-0418782.					700	GACGTGAACAGCCAGAGACCATGGAACCCGCTGGCGGCTGTCCAAATGGGCTGTATC	759
XX	07-APR-1995; 95US-0418782.					703	GSTAGCGGGATC-----TGGAGAACCCTGGCGCGGTGCGAGTGGGGCTGATC	753
XX	18-APR-1994; 94US-0228662.					760	FATGTGACCCGAGAGGTGTAAAGCCACCTGATGGCTGAGAACGGCACACGAGAGTA	819
PA	(MAYO-) MAYO FOUNDATION.					754	TAGGTGAACCGGAGGCGTGAACCGAACCGGACCCCTTGGCGCGGCTGCGACT	813
PI	Cockerill FR, Kline BC, Uhl JR;					820	CTTGAACCTTCGGCGGTATGGCGATGAACGAGCAAAAACCGACGCGCTCACAGCTGGC	879
DR	WPI; 1997-424226/39.					814	CGGAGAGGTTTCGGCGCATGGCGCATGAGAGCTGTAATGGGCTGTGAGAGCTGTTAGC	879
DR	P-PSDB; AWM6596.					880	GGCCACACCGCTGATTAATGTACAGGTATGGCAATGCTCTGGGTAGCGCTGAGCCA	939
PT	Determining susceptibility of Mycobacterium tuberculosis strains to Isoniazid - by detecting mutation(s) in the catalase-peroxidase gene, katG					874	GGTCACATTTTCGTGAAGACCATGGCGCGCGCGCGCTGCTGGTGGCGCGCGAACCC	933
XX	Claim 1; Column 31-36; 38pp; English.					940	AAAGCTCTGAGCTTGAACAGGGGTGAGTTGGGCGCAACCCCAACATGACGGGCAAG	999
XX	This DNA sequence comprises a consensus of the Mycobacterium tuberculosis wild-type katG (catalase-peroxidase) gene. Claimed methods of rapidly identifying strains of M. tuberculosis which are resistant to the tuberculostatic drug Isoniazid (INH) are based on the discovery of 4 mutations in the wild-type gene sequence that confer INH resistance and which coincidentally result in the addition or deletion of restriction endonuclease sites. Primers (see below) are provided for the PCR amplification of the katG gene of a test strain of Mycobacterium tuberculosis. The presence of a AAT50403-06) for use in the restriction fragment length polymorphism analysis of the amplified gene, and hence for determination the susceptibility to INH of the strain.					930	GCACACACCGCTGATTAATGTACAGGTATGGCAATGCTCTGGGTAGCGCTGAGCCA	999
XX	Sequence 2331 BP; 455 A; 734 C; 763 G; 379 T; 0 other;					1120	GTTGGCCACCATTTGGGACCGATGACATAAAAGGAAACACGCGGTGTAGCGGACG	1179
XX	Query Match 27.3%; Score 618.6; DB 18; Length 2331; Best Local Similarity 58.7%; Pred. No. 3.4e-165; Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;					1114	GGGCGCTTGGCAATACACGGCGAGGCGCGCGGTGCGGCACATCCCGGAGCCGCTTC	1173
XX						1170	GACCCCTTATTTGCCACACCGCATCATGACGATGGCGATTTGGCATTAAGTAAAT	1239
XX						1174	GGCGGGCG---AGGGGCTTCGCGACATGCTGCCACTGCTCTGCGGGTGGAT	1290
QY	160 AACAAAGATTGGTGGCGGAAGGTTGAACCTGGATATTTTGCATCCGACGATCGCAA	219				1240	CGACCTATCGGCTATCTGCGGAAATTCATGGCGGATCTGAGTACTTCAGAAAGCT	1290
Db	172 AACGAGGACTGGTGGCGCAACCGGCTCAATCTGAAGGTACTGCACCAACCGCGCGTC	231				1231	CGATCTATGAGCGGATACCGGTGCGGTGGCAACACCCCGAGAAATTTGGCGGACGAG	1290
QY	220 TCAGACCGGATGGATCCGGATTTCACTACCGTGAAGATACGACGCTGATTCGAC	279				1300	TTGCGAAGCGGTGGTGAAGTGAACCGCTGACCTGGCGCGGAAATCAGCTTACATC	1359
QY						1291	TTGCGAAGCGGTGGTGAAGTGAACCGCTGACCTGGCGCGGAAATCAGCTTACATC	1359

Qy 1360 GGCCTGGAGTGGCGGACAGAGCTGATTGGCAGACCCGATTCGCGAGGTATACACC 1419  
 Db 1361 GGCCTGGAGTGGCGGACAGAGCTGATTGGCAGACCCGATTCGCGAGGTATACACC 1410  
 Qy 1410 GACTACTGCGAGAGTGG- - - - -TCAAGCAGAAAATTCACAAAGTGGCGTG 1467  
 Db 1411 GACTACTGCGAGAGTGG- - - - -TCAAGCAGAAAATTCACAAAGTGGCGTG 1470  
 Qy 1468 AGCAATGAGAGTGGTTCACACCGCTGGCGAGACGTGCGGTACTGCGGTTCGAG 1527  
 Db 1469 AGCAATGAGAGTGGTTCACACCGCTGGCGAGACGTGCGGTACTGCGGTTCGAG 1530  
 Qy 1471 ACTGTCCTACAGCTAGTTCGACCGCATGGCGGCGTCTGCTTCGCTGTGAGGAC 1537  
 Db 1472 ACTGTCCTACAGCTAGTTCGACCGCATGGCGGCGTCTGCTTCGCTGTGAGGAC 1540  
 Qy 1528 ATGGCGGCGGTGCTAACGGTGGCGCGCATTCGCTTGGCGCCACAGACAGTGGCAGGCG 1587  
 Db 1529 ATGGCGGCGGTGCTAACGGTGGCGCGCATTCGCTTGGCGCCACAGACAGTGGCAGGCG 1590  
 Qy 1588 AACGAGCGGA- - -GCGCTCGGCGAAGTGTGAGGCTGTACGAGAGAGTCT- - - - - 1636  
 Db 1589 AACGAGCGGA- - -GCGCTCGGCGAAGTGTGAGGCTGTACGAGAGAGTCT- - - - - 1639  
 Qy 1591 AACGAGCGGAGCGGGATCTCGCAAGTCTATTCGCAACCTGGAAGAGATCAGGAGTCA 1650  
 Db 1592 AACGAGCGGAGCGGGATCTCGCAAGTCTATTCGCAACCTGGAAGAGATCAGGAGTCA 1653  
 Qy 1637 - - - - -CTCCGTCACACGCGCGCTAGATCGCGGAGCGTATCTTTCGGCG 1680  
 Db 1638 - - - - -CTCCGTCACACGCGCGCTAGATCGCGGAGCGTATCTTTCGGCG 1683  
 Qy 1681 GGTAGCTAGCAGTCCGGAAGCGGAGACGAGTGTACGAGTTCGCGGTTCGCTG 1740  
 Db 1682 GGTAGCTAGCAGTCCGGAAGCGGAGACGAGTGTACGAGTTCGCGGTTCGCTG 1743  
 Qy 1711 GCGTGTGCGCCATAGAGAAAGCAGAAAGCGGCTGGCCACACATACGCTGCGCTTC 1770  
 Db 1712 GCGTGTGCGCCATAGAGAAAGCAGAAAGCGGCTGGCCACACATACGCTGCGCTTC 1773  
 Qy 1741 CTGAAAGCGCGTGGCGATCGACCGCGAGATGACCGACGAGATCTCTTCGACCGCGTG 1800  
 Db 1742 CTGAAAGCGCGTGGCGATCGACCGCGAGATGACCGACGAGATCTCTTCGACCGCGTG 1803  
 Qy 1771 ACCCGGGCGCGCGAGTGTGCGAGTTCGAGTACGAGTACGAGTTCGCGGTGCGTG 1830  
 Db 1772 ACCCGGGCGCGCGAGTGTGCGAGTTCGAGTACGAGTACGAGTTCGCGGTGCGTG 1833  
 Qy 1801 GAGCGCGTGGCGGATGCTCCGCACTGCGAGTTCGAGTACGAGTTCGCGGTGCGTG 1860  
 Db 1802 GAGCGCGTGGCGGATGCTCCGCACTGCGAGTTCGAGTACGAGTTCGCGGTGCGTG 1863  
 Qy 1831 GAGCGCGTGGCGGATGCTCCGCACTGCGAGTTCGAGTACGAGTTCGCGGTGCGTG 1890  
 Db 1832 GAGCGCGTGGCGGATGCTCCGCACTGCGAGTTCGAGTACGAGTTCGCGGTGCGTG 1893  
 Qy 1861 GAGTGTGCTGATCTGCGCGAGCTGATGGGCTTAACCGCGCGGAAATGACCGTGGTG 1920  
 Db 1862 GAGTGTGCTGATCTGCGCGAGCTGATGGGCTTAACCGCGCGGAAATGACCGTGGTG 1923  
 Qy 1891 TACATGCTGCTGCGACGAGCGACCTGTAGCTGCTGTGCTGCGACCTGCTGCTGGGT 1950  
 Db 1892 TACATGCTGCTGCGACGAGCGACCTGTAGCTGCTGTGCTGCGACCTGCTGCTGGGT 1953  
 Qy 1921 CTGGGGTGTATGGCGTACTTGGCGACCACTATGTTGGCACCACCAACGCGGTATCAC 1980  
 Db 1922 CTGGGGTGTATGGCGTACTTGGCGACCACTATGTTGGCACCACCAACGCGGTATCAC 1983  
 Qy 1951 GTAGGTGCGCTTGGCGTCTTGGCGCACTTACGAGCGCTTACGCTGGGCGTTCACG 2010  
 Db 1952 GTAGGTGCGCTTGGCGTCTTGGCGCACTTACGAGCGCTTACGCTGGGCGTTCACG 2013  
 Qy 1981 GATTGTGAAGCGCATTTGACACAGACTTTTGTGACCTGACCGATATGGGACAGC 2040  
 Db 1982 GATTGTGAAGCGCATTTGACACAGACTTTTGTGACCTGACCGATATGGGACAGC 2043  
 Qy 2011 GAGGCTCGGAGTACTATACACAGCTTCTGAGCTGCTGAGCTGAGTGGTATCAC 2070  
 Db 2012 GAGGCTCGGAGTACTATACACAGCTTCTGAGCTGCTGAGCTGAGTGGTATCAC 2073  
 Qy 2041 TGGAGC- - -CGTATGATGACAGCTCTACGAATCGCGCGACGACGCGGCGGTG 2097  
 Db 2042 TGGAGC- - -CGTATGATGACAGCTCTACGAATCGCGCGACGACGCGGCGGTG 2100  
 Qy 2071 TGGAGCGCTTGGCGAGAGTACGCGGACCTTACGAGGCAAGATGCGAGTGGCAAGTG 2130  
 Db 2072 TGGAGCGCTTGGCGAGAGTACGCGGACCTTACGAGGCAAGATGCGAGTGGCAAGTG 2133  
 Qy 2098 AATGTCAGCGCTTGGCGGAGTATGATATTGTTGCTCACTGCTACTGCGCTTTAC 2157  
 Db 2099 AATGTCAGCGCTTGGCGGAGTATGATATTGTTGCTCACTGCTACTGCGCTTTAC 2160  
 Qy 2131 AATGTCAGCGCTTGGCGGAGTATGATATTGTTGCTCACTGCTACTGCGCTTTAC 2190  
 Db 2132 AATGTCAGCGCTTGGCGGAGTATGATATTGTTGCTCACTGCTACTGCGCTTTAC 2193  
 Qy 2158 GCAGATGATACCGCGAGAGAGTTCGTCAGAGACTTCTGCGCGCG 2217  
 Db 2159 GCAGATGATACCGCGAGAGAGTTCGTCAGAGACTTCTGCGCGCG 2220  
 Qy 2191 GTCCGGGTATGGGCGGATGACGCGACGCGAGGATTCGTGAGGACTGCTGCTGCTGCC 2250  
 Db 2192 GTCCGGGTATGGGCGGATGACGCGACGCGAGGATTCGTGAGGACTGCTGCTGCTGCC 2253  
 Qy 2218 TGGACAAAGTATGACGCGACGCTTTGACGT 2252  
 Db 2219 TGGACAAAGTATGACGCGACGCTTTGACGT 2255  
 Qy 2251 TGGACAAAGTATGACGCGACGCTTTGACGT 2285  
 Db 2252 TGGACAAAGTATGACGCGACGCTTTGACGT 2288

RESULT 9  
 AAX00817  
 ID AAX00817 standard; DNA; 2331 BP.  
 XX  
 AX  
 AX00817;  
 XX  
 DT 26-MAR-1999 (first entry)

M. tuberculosis catalase peroxide (katG) gene sequence.  
 Catalase-peroxide; katG; isonicotinic acid hydrazide; tuberculosis;  
 INH; diagnosis; detection; as.  
 Mycobacterium tuberculosis.  
 Key Location/Qualifiers  
 CUS 70..2292  
 /name= "katG"  
 /gene= "katG"  
 /product= "catalase peroxide"  
 /note= "the start codon is not indicated"

W09850585-A1.  
 12-NOV-1998.  
 06-MAY-1998; 98W0-US09285.  
 07-MAY-1997; 97U5-0852219.  
 (MAYO-) MAYO FOUNDATION.  
 Cockerill FR, Kline BC, Uhl JR;  
 WPI: 1995-070099/06.  
 P-PSDB: AAW95398.

Detection of Mycobacterium tuberculosis - by amplifying katG gene  
 and detecting specific fragment, and optionally identifying  
 INH-resistant strains by detecting specific mutation  
 Claim 3; Fig 7; 83pp; English.

The invention relates to a novel method of detecting Mycobacterium  
 tuberculosis. The method comprises amplifying the DNA in the samples to  
 generate a detectable amount of amplified DNA comprising a catalase-  
 peroxide (katG) DNA fragment with sequence of bases 904-1523 of the  
 katG gene. The method further comprises determining the method optionally  
 further comprises determining if the katG gene sequence is the same as  
 the threonine mutation in codon 315 (S315T mutation), indicative of an  
 isonicotinic acid hydrazide (INH)-resistant M. tuberculosis strain. The  
 method can be used to detect M. tuberculosis in biological fluids,  
 especially human sputum, useful to diagnose tuberculosis. This disease is  
 a major cause of human morbidity and mortality, and conclusive diagnosis  
 and subsequent treatment depends on identification of the etiologic agent  
 and subsequent treatment. The method allows for the identification of  
 INH-resistant strains have emerged. The method allows such drug-resistant  
 strains to be identified. The present sequence represents the wild-type  
 M. tuberculosis katG gene sequence.

Sequence 2331 BP; 455 A; 734 C; 763 G; 379 T; 0 other;  
 Query Match 27.3%; Score 618.6; DB 20; Length 2331;  
 Best Local Similarity 58.7%; Fragment 46.6%;  
 Matches 1253; Conservative 0; Mismatches 819; Indels 63; Gaps 8;

Qy 160 AACAAAGATGTGGTGGCGGAGGTTGAACTGGATATTTTCATCAGACGATGCGAAA 219  
 Db 172 AACACGAGTGTGGCGGCAACCGGCTCACTCTGAAGTACTGACACAAACCGCGCTC 231  
 Qy 220 TCAGACCGATTCGCGGATTCACCTACCGCTGAAGACGACGATGATTCGAC 279  
 Db 232 GCTGACCGATGGGTGGCGCTTGACCTTGGCCGCGAGTTCGACCATCGCTGAC 291  
 Qy 280 GCGCTGAAGAAAGATGTCACCGCGTGTGATGACCATGACCAAGATGTTGGCGCCGTCAC 339  
 Db 292 GCGCTGAAGAAAGATGTCACCGCGTGTGATGACCATGACCAAGATGTTGGCGCCGTCAC 351  
 Qy 340 TGGGCGGACCTACCGCGGTTGATGATGATGCTGTGGCATCTCCGCTGGGACCTACCGGT 399  
 Db 341 TGGGCGGACCTACCGCGGTTGATGATGATGCTGTGGCATCTCCGCTGGGACCTACCGGT 399

[illegible][illegible]

RESULT 10
ABX09140/C
ID ABX09
XX
AC ABX09
XX
DT 08-AP
XX
DE Mycob
XX
KW Mycob
KW mycob
XX
OS Mycob
XX

OS *Mycobacterium tuberculosis*.





Dd 38857 GCTGTGTCGCCCATATGAGAAAGACGAAAGAGGGGCTGGCCACACATACACGCTGCGCCTTC 38798  
 QY 1741 CTGAAAGAGCCGCGGATGCGGACCCGCGAGATGACGACGACGATGCTTCCGACCCGTC 1800  
 Dd 38797 ACCCGGGCGGACGGATGCGTCCGAGGAACAAACGACGATGATGCTTTGGCGCTGTC 38738  
 QY 1801 GAGCCGCTGGCGCATGGCTTCGCGCACTGGCAGAGAAAGATATGGTGAAAGCCGAA 1860  
 Dd 38737 GAGCCGAGGAGATGAGCTTCCGAACTACTCTGGAAAGGGGACACCGCTTCCGCGCGAG 38678  
 QY 1861 GAGATGCTGCTGAGATGCGGACGATGATGGCTTAACCGCGCGCGGAAATGACCTGTC 1920  
 Dd 38677 TACATGCTGCTGCAAGGACGAACCTGCTTACGCTCACTGCGCCCTGAGATGAGGCTG 38618  
 QY 1921 CTGGGCGGTATCCCGCTATGCGGACCACTATGCTGGCACAACAGCGCGTATTCCAC 1980  
 Dd 38617 GTAGGCTGCGGCTTCTGCGCGCAACTACACAGCCCTTACCGCTGGGCGGTTCAC 38558  
 QY 1981 GATTTGGAGGCTGATGACCAACGACTTTTGTGACACTGACGAGCTTACCGCTGGGCGGTTCAC 38548  
 Dd 38557 GAGGCTCCGAGTCACTGACCAACGACTTTGTGACACTGCTGACATGGATTCAC 38498  
 QY 2041 TGGAGC---CGTAGTAGAGAACGCTACGAATTCGCGACCCGACAGCGCTGCGCTG 2097  
 Dd 38497 TGGAGGCTTCCGACAGATGAGGCTACGAGGAGAGATGAGTGGAGAGTG 38438  
 QY 2098 AAGTGGACCGCTCGCGGTGATCTGGTATGGTTCACACTGCTACTCGCTCTTAC 2157  
 Dd 38437 AAGTGGACCGCGCGGTGACCTGGTCTTGGGTCCACTGGAGTTCGCGCGCT 38378  
 QY 2158 GAGAGATGATGAGGAGAGAGTACGCGGAGAGTTCGTCAGAGCTTCGTCGCGCG 2217  
 Dd 38377 GTGGGCTTGTGGCGGATGAGCGGACGCTGCGAGCTTCGTGCTGCTG 38318  
 QY 2218 TGGAGCAAGTATGAGGCGGACGCTTTCGAGCT 2252  
 Dd 38317 TGGGCAAGGTGATGACCTGACAGGTTTCGAGCT 38283  
 RESULT 11  
 AAI99683/C  
 ID AAI99683 standard; DNA; 4403765 BP.  
 AC AAI99683:  
 XX  
 XX 15-JAN-2002 (first entry)  
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
 XX  
 XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
 XX  
 XX Mycobacterium tuberculosis.  
 OS  
 PN US6294328-B1.  
 XX  
 XX 25-SEP-2001.  
 PD  
 XX  
 XX 24-JUN-1998; 98US-0103840.  
 XX  
 XX 24-JUN-1998; 98US-0103840.  
 PR  
 XX (GENO-) INST GENOMIC RES.  
 DX  
 XX FLEISCHMANN RD, White OR, Fraser CM, Venter JC;  
 PI  
 XX NFI; 2001-647261/74.  
 DR  
 XX  
 XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ.  
 XX

PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
 XX The invention relates to evaluating strain variation within and between  
 CC different Mycobacterium tuberculosis pathogen, including the  
 CC Mycobacterium tuberculosis or related Mycobacterium species, by  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
 CC H37Rv (AAI99688). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.  
 XX  
 SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other:  
 Query Watch 27 38; Score 618.6; DP 22; Length 4403765;  
 Best Local Similarity 58 74; Positives 1,265; Indels 63; Gaps 8;  
 Matches 1253; Conservative 0; Mismatches 815;  
 QY 160 AACAAAGATTTGGTGGCGGAAAGGTTTGACCTGTGATATTTCATCAGCAAGATCGCAAA 219  
 Dd 2153267 AACCAAGATTTGGTGGCGGAAAGGTTTGACCTGTGATATTTCATCAGCAAGATCGCAAA 2193268  
 QY 220 TCAGACCGATGATGCGGATTTGACCTGTGATATTTCATCAGCAAGATCGCAAA 2193269  
 Dd 2153207 GCTGACCGATGGTGGCGGTTTCCACTATGCGGAGGTGGGACCATGACGCTTGAC 2153148  
 QY 280 GCGCTGAGAAAGATGTTCCAGCGGTTTGATGACGATAGCCAGAGTGTGGCGCGCTGAC 339  
 Dd 2153147 GCGCTGAGCGGCGGACATGAGGAAGTGTGACACCTGCGGCGGCTGTGGCGCGCGAC 2153088  
 QY 340 TGGGCGCACTAGCGCGGTTGATGACGATGAGTGGGCTGCGGCTGCGGCGCGCTACCT 399  
 Dd 2153087 TACGGCACTAGCGCGGCTGTTTATCGGATGGCTGGCGGCTGCGGCGCGCTACCT 2153028  
 QY 400 ATTGCTGATGGCGCTGGGCGGCTGTGACGGAAGCCAGCGGTTTGACCGCTCAACCTC 459  
 Dd 2153027 ATGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCTTACGCGCTTAAACG 2152968  
 QY 460 TGGCGGACAGCTGAGCTGGAATAAGCGCGGCTGCTGTGCGGCTGATGANGAAG 519  
 Dd 2152967 TGGCGGACAGAGCTGGAATAAGCGCGGCTGCTGTGCGGCTGATGANGAAG 2152908  
 QY 520 TACGCAACAAATCAGCTTGGCGGACAGCTGATGATTCGCTGGCGCGCTGTATGAG 579  
 Dd 2152907 TACGCAACAAATCAGCTTGGCGGACAGCTGATGATTCGCTGGCGCGCTGTATGAG 2152848  
 QY 580 TCACTGAGCTTACCTGCTTACGCGCTCTGTTTGGCGCGCTCGATATTTCGCAACCCGAA 639  
 Dd 2152847 TCGATGGCTTCAAGAGCTTTCGCGCTTGGCGCTGCGCGGCTGCAAGAGCGG 2152790  
 QY 640 AAGATATCTACTGGGCTGAGGAAAGAGTGGCTGGACCTTCTGAGCAAGCTACGCG 699  
 Dd 2152789 ATGAGGCTTATTTGGGCGAAGAGCCCTTGGCTG-----CGCTGAGCGCTTACG 2152737  
 QY 700 GAGGTGAACAACCCAGAGACCATGGAACCCGCTGGCGGCTGTCCAAATGGCTGCTATC 759  
 Dd 2152736 GGTGAGCGGATC-----TGGGAACCCGCTGGCGGCTGCAATGGCTGCTATC 2152686  
 QY 760 TATGTGAACCCGAGGTGTTTACGCGCACTGATCCGCTGAGACCGGACAGCGATGA 819  
 Dd 2152685 TATGTGAACCCGAGGTGTTTACGCGCACTGATCCGCTGAGACCGGACAGCGATGA 2152626  
 QY 820 CTGGAACCTTCGCGCTGATGCGGATGACGAGGAAAGAGCCGCGCTTCACAGCTGCG 879  
 Dd 2152625 CGCGAGAGTTTCGCGGCGATGACCATTAACGAGGTGGAACAGAGCGGCTGATCTCGCG 2152566  
 QY 880 GCGCAACCGCTCGGTATTGTGACGGTAATGGCTGCGGCTTACGCGCTGACCGCA 939  
 Dd 2152565 GGTGACACTTTCGCTTACGACCATGCGCGGCGGCGGCTGCTGCTGCGCGCGACCC 2152506

QY 940 AAAGGCTCTCAGCTGTGAACACAGGCGTTCAGTGGGCAACCCCACTGACAGGCGAG 999  
 Db 2152505 GAGGCTGCTCCGCTGGAGCAGATGGCGTGGCTGGGAAGAGCTGTATGCGACGGAAAC 2152446  
 QY 1000 GCAAGCAACGCGGTGACTCGGCTATGGAAGGTGCTTGGACACCAACCCAGGAATTC 1059  
 Db 2152445 GTATGAGCAAGCTCACAGCGGCGATCGAGCTGTATGGACCAACCGAGGAATG 2152386  
 QY 1060 GATATGGGCTATTCCTACCTGCTGCTGCTACAACTGCGACACTGGAAGAGTGTGGC 1119  
 Db 2152385 GACAACTGTTCTCTCAGAGTCTGTATGCGCTACGAGTGGGAGCTGACGAGAGCCCTGCT 2152326  
 QY 1120 GTGGCCACAGCTTGGGACACGATTTGACATMAAAGGAAACAAAGCGGTGTGACGCCAGC 1179  
 Db 2152325 GGCGCTGTGGCAATACACGCTCAAGAGAGGCGCGCGTGGCGGACCACTACCCGACCCCTTC 2152266  
 QY 1180 GACCTCTATTCGACCAACCCGATGATGACAGTGGGATATGCGCATTAAGAGTAAAT 1239  
 Db 2152265 GCGGGGCGC---AGGGGCGTCCCGACGACTGCTGGCCACTGACCTCTGCTGGCGGTGGAT 2152209  
 QY 1240 CGACACTATCGGCTATTCCTCGCAAAAATTCATGCGCGGATCTGTGACTTCAAGAAAACCT 1299  
 Db 2152208 CGGATCTATGACGGGATCAACCGCTGCTGGCTGACMACCCCGAGGAATTGGCGCGAGGAC 2152149  
 QY 1300 TTGCGAGGAGCTGTACAGTGTACGACGACGAGCTGGCGGCGGAATCATGACGATCTC 1359  
 Db 2152148 TTGCGGACGCTGTGTACAGCTGATCATCCGAGAGACTGGGTCCGCTTGGAGATACCTT 2152089  
 QY 1360 GCGCGGAAGTGGCGGAGAGACCTGATTGGCAAGACCCGATTCGGCGCATACAC 1419  
 Db 2152088 GGGCGCTGTGCTCCGANGAGACCCCTGCTGTGGAGATCCGGTCCGTGGCTGACGAC 2152029  
 QY 1420 GACTACTCGCAAGAGTGG-----TCAAGCAGAAATTCGACAAAGTGGCGCTG 1467  
 Db 2152028 GACTCTGTGGCGAAGCGAGATGTCGACGCTTAAAGCGCAGATCGGCGATCGGGATG 2151969  
 QY 1468 AGCATATGATGAGTGTCTCCACCGCTTGGGACAGTGGCGCTACTTATTCGCGTTCCGAT 1527  
 Db 2151968 ACTGCTCTACACTAGTATTCGACAGCATGGCGCGCGCTGCTGTGCTGTGATGCGAC 2151909  
 QY 1528 ATGCGCGCGGTGCTAGTGGTGGCGGCTTGGCGCCACAGACGAGTGGCGAGGCG 1587  
 Db 2151908 AAGCGCGCGGCGGCGCAACGSGTGTGGCTCGCGCTCGACGACAAAGTGGGTGGGAGTCT 2151849  
 QY 1588 AAGCGCGCGA---GGCGCTGGCGAAGTGTGAGCGGTCTACGACAGATCT----- 1636  
 Db 2151848 AAGCGCGCGAGTGTCCGAGTGTCTGCGAGTATGCGACCTCGAGAGATCGAGAGTCA 2151789  
 QY 1637 -----CTGGCGACACCGCGTAGCATGTCGCGACCTGATGCTTCCTGGCG 1680  
 Db 2151788 TTCACCTCGCGCGCGCGGCGAACAATCAAGTGTCTTCGCGACACTGCTGCTCGCTG 2151729  
 QY 1661 GTATGCTAGGAGTGTGAGAAAGCCGCGAAGCGACAGGTATGAGATGGCGTTCCTCTC 1740  
 Db 2151728 GCTGTGCTGCCCTAGAGAGACCGAGAGGCGTGGCCACAGCTCTGCGCGGCTTC 2151669  
 QY 1741 GTAAAGAGCTGTGGATGCGACCGCGAGATGACCGACAGACTCTTCGCGACCGCTG 1800  
 Db 2151668 ACCCGCGCGCGCGAGGATGCTCGCAGGACAAACACGATGGAATCTTTCGCGCTGTG 2151609  
 QY 1801 GAGCGCTGGCGCATGGCTTCGCGCACTGGCAGAGAAAGATGTGTGGTGAAGCGGAA 1860  
 Db 2151608 GAGCGCGAGTGTGGATGCGCAGCTGATGGCTTACCGGCGCGGCGGAAATGACCTGCTG 2151549  
 QY 1861 GAGATGCTGTGATGTGGCAGCTGATGGCTTACCGGCGCGGCGGAAATGACCTGCTG 1920  
 Db 2151548 TACATGCTGTGCGAAGCGCAACCTGCTTACGCTCAGTGCCTCGATGATGAGTGGCTG 2151489  
 QY 1921 CTGGGCGGTATGCGGCTATGGCGCACCACTATGTCGCGACCAACAGCGGCTATTCACC 1980  
 Db 2151488 GTAGTGTGCCCTGGCGGCTCTGGCGGCACTACAGGCGCTTACGCGTGGCGGTGTCCAC 2151429

QY 1981 GATTGTGAAGCGCGATTTGACCAACGACTTTTTGTGAACCTGACGATATGGGGAACAGC 2040  
 Db 2151428 GAGCGCTCCGAGTCACTGACCAACGACTTCTTCGTGACCTGCTGCAATGGGTATCAC 2151369  
 QY 2041 TGAAGC---CGTAGTGTAGCAACGCTTACGAAATCCGCGACCGCAAGACCGTGGCGTG 2097  
 Db 2151368 TGGGAGCGCTTCGCCAGCAGATGACGGGAGCTTACCAGGGCAAGGATGCGCATGAAGTG 2151309  
 QY 2098 AAGTGGACCGCTTCGCGGTGATGCTGTATTTGGTTCCAACTCGCTACTCGCTCTTAC 2157  
 Db 2151308 AAGTGGACCGCGCGCGGTGACCTGTGTTTCGGGTCCACTCGAGTGTGGGCGCTT 2151249  
 QY 2158 GCGAAGAGTGTAGCCGACGAGATAACGAGAGTGTCTGAGAGCTTTCTGCGCGCGC 2217  
 Db 2151248 GTTCGAGTCTATGCGCGCGATGACGCGGACCGCAAGTTCTGCGAGCACTTTCGCTGCC 2151189  
 QY 2218 TGGACCAAGTGTGTAACCGCGACCGCTTTCGAGCT 2252  
 Db 2151188 TGGGAGAGGTGTATGTAACCTCTGCAAGCTTGCAGCT 2151154

## RESULT 12

AA199682/C

ID AA199682 standard; DNA; 4411529 BP.

XX

XX AA199682;

XX

DT 15-JAN-2002 (first entry)

XX

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.

XX

XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 XX variation; epidemiology; patient treatment; epidemic monitoring; ds.

XX

OS Mycobacterium tuberculosis.

XX

PN US6294328-B1.

XX

PD 25-SEP-2001.

XX

XX 24-JUN-1998; 9808-0103840.

XX

PR 24-JUN-1998; 9808-0103840.

XX

XX (GENO-) INST GENOMIC RES.

XX

XX Fleischmann RD, White OR, Fraser CM, Venter JC;

PT

XX WPI; 2001-647261/74.

XX

PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ .

XX

XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.

XX

CC The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculois bacterial pathogen,  
 CC Mycobacterium tuberculosis or related mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions that differ in the  
 CC sequence of the genome that correspond to positions in the complete  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199682) and  
 CC H37Rv (AA199682). The method is useful for evaluating strain variation of  
 CC Mycobacterium tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.

CC

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at [seqdata.uspto.gov/sequence.html?docID=6294328B1](http://seqdata.uspto.gov/sequence.html?docID=6294328B1).

XX

XX Sequence 4411529 BP; 758555 A; 1449983 C; 1444602 G; 758379 T; 0 other;















**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 12:53:54 ; Search time 578.569 Seconds  
100000 alignments  
101447.786 Million cell updates/sec

Title: US-09-884-889-5

Perfect score: 2262  
Sequence: 1 atgaataacgcattcgctga.....gtttcagctgcgtgtaa 2262

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1731409 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA.\*

1: /cgn2.6/ptodata/2/pubna/US07\_PUBCOMB.seq.\*  
2: /cgn2.6/ptodata/2/pubna/US06\_PUBCOMB.seq.\*  
3: /cgn2.6/ptodata/2/pubna/US06\_PUBCOMB.seq.\*  
4: /cgn2.6/ptodata/2/pubna/US06\_PUBCOMB.seq.\*  
5: /cgn2.6/ptodata/2/pubna/US07\_NEW\_PUB.seq.\*  
6: /cgn2.6/ptodata/2/pubna/US07\_NEW\_PUB.seq.\*  
7: /cgn2.6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*  
8: /cgn2.6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*  
9: /cgn2.6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*  
10: /cgn2.6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*  
11: /cgn2.6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*  
12: /cgn2.6/ptodata/2/pubna/US09\_PUBCOMB.seq.\*  
13: /cgn2.6/ptodata/2/pubna/US10\_PUBCOMB.seq.\*  
14: /cgn2.6/ptodata/2/pubna/US10\_PUBCOMB.seq.\*  
15: /cgn2.6/ptodata/2/pubna/US10\_NEW\_PUB.seq.\*  
16: /cgn2.6/ptodata/2/pubna/US10\_NEW\_PUB.seq.\*  
17: /cgn2.6/ptodata/2/pubna/US10\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2262	100.0	2262	10	US-09-884-889-5
2	672.2	29.7	2238	10	US-09-884-889-7
3	618.6	27.3	82993	15	US-10-080-170-645
4	137.2	6.1	620	11	US-09-940-925A-144
5	137.2	6.1	620	11	US-09-940-925A-145
6	137.2	6.1	620	11	US-09-940-925A-146
7	137.2	6.1	620	11	US-09-940-925A-147
8	137.2	6.1	620	11	US-09-940-925A-148
9	137.2	6.1	620	11	US-09-941-193A-145
10	137.2	6.1	620	11	US-09-941-193A-146
11	137.2	6.1	620	11	US-09-941-193A-147
12	135.6	6.0	620	11	US-09-940-925A-143
13	135.6	6.0	620	11	US-09-940-925A-144
14	135.6	6.0	620	11	US-09-940-925A-145
15	135.6	6.0	620	11	US-09-940-925A-146
16	135.6	6.0	620	11	US-09-941-193A-143

17	135.6	6.0	620	11	US-09-941-193A-145
18	135.6 <td>6.0 <td>620 <td>11</td> <td>US-09-941-193A-147</td> </td></td>	6.0 <td>620 <td>11</td> <td>US-09-941-193A-147</td> </td>	620 <td>11</td> <td>US-09-941-193A-147</td>	11	US-09-941-193A-147
19	135.6 <td>6.0 <td>620 <td>11</td> <td>US-09-941-193A-149</td> </td></td>	6.0 <td>620 <td>11</td> <td>US-09-941-193A-149</td> </td>	620 <td>11</td> <td>US-09-941-193A-149</td>	11	US-09-941-193A-149
20	91	4.0	391	10	US-09-825-574-4
21	91	4.0	391	10	US-09-882-945A-4
22	89.4	4.0	391	10	US-09-825-574-1
23	89.4	4.0	391	10	US-09-825-574-2
24	89.4	4.0	391	12	US-09-882-945A-1
25	89.4	4.0	391	12	US-09-882-945A-2
26	87.8	3.6	391	12	US-09-882-945A-3
27	87.8	3.6	391	12	US-09-882-945A-3
28	42	1.9	2802	9	US-09-815-243-7720
29	42	1.9	30013	10	US-09-764-877-3297
30	41.6	1.8	516	14	US-10-156-761-4202
31	41.6	1.8	9025608	14	US-10-156-761-1
32	41.4	1.8	671	14	US-10-184-634-346
33	41.4	1.8	971	14	US-10-184-634-346
34	40.8	1.8	1041	14	US-10-156-761-4102
35	40.8	1.8	1041	14	US-10-156-761-4102
36	40.8	1.8	2379	9	US-09-815-243-7868
37	40.8	1.8	2379	14	US-10-127-032-80
38	40.8	1.8	9025608	14	US-10-156-761-1
39	40.4	1.8	3206	14	US-10-156-761-6845
40	40.2	1.8	1504	14	US-09-815-243-7884
41	39.6	1.8	1844	9	US-09-815-243-7884
42	39.6	1.8	553	14	US-10-184-634-402
43	39.2	1.7	88421	10	US-09-976-059-1
44	39.2	1.7	88421	10	US-10-084-843-139
45	39	1.7	815	12	US-10-084-843-139

## ALIGNMENTS

## RESULT 1

US-09-884-889-5  
; Sequence 5, Application US/09884889  
; Patent No. US20020102680A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: SANYAL, Indrajit  
; APPLICANT: ADHIKARI, Robert  
; TITLE OF INVENTION: CATALASES  
; FILE REFERENCE: DIVER1100-4  
; CURRENT APPLICATION NUMBER: US/09/884,889  
; CURRENT FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 09/412,347  
; PRIOR FILING DATE: 1997-10-16  
; PRIOR APPLICATION NUMBER: US 08/951,844  
; PRIOR FILING DATE: 1997-10-16  
; PRIOR APPLICATION NUMBER: US 08/674,887  
; PRIOR FILING DATE: 1996-07-03  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; TYPE: DNA  
; ORGANISM: Alcaligenes (Deleya) aquamarinus  
US-09-884-889-5

Query Match 100.0%; Score 2262; DB 10; Length 2262;  
Beat Local Similarity 100.0%; Pred. No. 0;  
Matches 2262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAATAACGCATTCGTCAGCATCTACACAGTACGTCGACAGACGATTCGACAGCATTC 60  
DB 1 ATGAATAACGCATTCGTCAGCATCTACACAGTACGTCGACAGACGATTCGACAGCATTC 60  
QY 61 GTTCCTTGGTATGCCCAAGGCATACAGCAATAAGGAGAGCTATGAGCGGTAATGT 120  
DB 61 GTTCCTTGGTATGCCCAAGGCATACAGCAATAAGGAGAGCTATGAGCGGTAATGT 120

QY 121 CCGTCTGTCAGCGGTGGTAAAGCCTTCAGACCGGTACTTCCAAACAGATTTGGTGGCCGAA 180  
 Db 121 CCGTCTGTCAGCGGTGGTAAAGCCTTCAGACCGGTACTTCCAAACAGATTTGGTGGCCGAA 180  
 QY 181 GGGTGAACCTGGAATTTGGTCAATCAGCAAGATGCAAAATAGACACCGATGGATCCGAT 240  
 Db 181 GGGTGAACCTGGAATTTGGTCAATCAGCAAGATGCAAAATAGACACCGATGGATCCGAT 240  
 QY 241 TTGCACTACCGGTGAAGAGTACACCAAGCTCGATTTGACGCGCTGGAAGAAATGTCAC 300  
 Db 241 TTGCACTACCGGTGAAGAGTACACCAAGCTCGATTTGACGCGCTGGAAGAAATGTCAC 300  
 QY 301 CGGTGATGATACCGATAGCAAGAGTGGTGGCCGCTGACTGGGCACTACGGCGGTTTG 360  
 Db 301 CGGTGATGATACCGATAGCAAGAGTGGTGGCCGCTGACTGGGCACTACGGCGGTTTG 360  
 QY 361 ATGATCGATATGGTCTGGACACTCCGCTGGACACTACGCTATTTGCTGATGGCGTGGGGC 420  
 Db 361 ATGATCGATATGGTCTGGACACTCCGCTGGACACTACGCTATTTGCTGATGGCGTGGGGC 420  
 QY 421 GGTGGTACCGGAAGCAGCGCTTTCGACACTCACTCTGGCGGACACAGCTCAGCTG 480  
 Db 421 GGTGGTACCGGAAGCAGCGCTTTCGACACTCACTCTGGCGGACACAGCTCAGCTG 480  
 QY 481 GATAAAGGCGCGCTCTGCTGGCGGATCAAGAGAGTACGCAAAATACAGCTGG 540  
 Db 481 GATAAAGGCGCGCTCTGCTGGCGGATCAAGAGAGTACGCAAAATACAGCTGG 540  
 QY 541 CCAGACTGATGATTTCTGGCTGGCAACCGTGGCTTATGATGTCATGGGCTTACCTGCTTAC 600  
 Db 541 CCAGACTGATGATTTCTGGCTGGCAACCGTGGCTTATGATGTCATGGGCTTACCTGCTTAC 600  
 QY 601 GCGTCTCTCTGGCGCGCTGCTGCTGGCGGATTTGGCAACCGCAAAAGATATCTAGTGGGTGAC 660  
 Db 601 GCGTCTCTCTGGCGCGCTGCTGCTGGCGGATTTGGCAACCGCAAAAGATATCTAGTGGGTGAC 660  
 QY 721 ATGGAACACCCGCTGGCGGCTGTCACCAATGGGCTGATCTATGTGAACCCGGAGGTGT 780  
 Db 721 ATGGAACACCCGCTGGCGGCTGTCACCAATGGGCTGATCTATGTGAACCCGGAGGTGT 780  
 QY 781 ACGGCCACCCCTGATCCGCTGAGAACCGCACAGAGTACTTGAACACTTCGCCCGCTATG 840  
 Db 781 ACGGCCACCCCTGATCCGCTGAGAACCGCACAGAGTACTTGAACACTTCGCCCGCTATG 840  
 QY 841 CCGATGACCAACCAACCCGCTGTCACAGCTGGGCGCAACAGCTGGCTGATTTG 900  
 Db 841 CCGATGACCAACCAACCCGCTGTCACAGCTGGGCGCAACAGCTGGCTGATTTG 900  
 QY 901 CAGGCTATAGGCAATGCTCTGGGTAGCCCTGACCCCAAGAAAGCTCTGACGCTGGAAC 960  
 Db 901 CAGGCTATAGGCAATGCTCTGGGTAGCCCTGACCCCAAGAAAGCTCTGACGCTGGAAC 960  
 QY 961 CAGGCGCTTAGGTTGGGCAACCCACATGACAGGCGCAGGCAAGCAACCGCGTGAACCTG 1020  
 Db 961 CAGGCGCTTAGGTTGGGCAACCCACATGACAGGCGCAGGCAAGCAACCGCGTGAACCTG 1020  
 QY 1021 GGTATGAGAGTGGTGGACACACCCACAGAAATGATATGATGGGCTATTTGACAGCTG 1080  
 Db 1021 GGTATGAGAGTGGTGGACACACCCACAGAAATGATATGATGGGCTATTTGACAGCTG 1080  
 QY 1081 CTGTTCGGCTACAAATTTGGGAAGCTGAAAAGAGTCCCTGGGTGCCCAACCAATTTGGGAACG 1140  
 Db 1081 CTGTTCGGCTACAAATTTGGGAAGCTGAAAAGAGTCCCTGGGTGCCCAACCAATTTGGGAACG 1140  
 QY 1141 ATGATCATMAAGGAACAAAGCGGTTGACGCGAGGACCCCTCTATTTGCGCAAC 1200  
 Db 1141 ATGATCATMAAGGAACAAAGCGGTTGACGCGAGGACCCCTCTATTTGCGCAAC 1200  
 QY 1201 CCGATCATACCGATGCGGATATGCGCGATAAGGTAAATCCGACCTATCCGCGCTATCTGCG 1260

1201 CCGATCATACCGATGCGGATATGCGCGATAAGGTAAATCCGACCTATCCGCGCTATCTGCG 1260  
 QY 1261 GAAAAATTCATCGCGGATCTGAGTACTTCTCAAGAAAGTTCGCGAAAGCGGTGTCAG 1320  
 Db 1261 GAAAAATTCATCGCGGATCTGAGTACTTCTCAAGAAAGTTCGCGAAAGCGGTGTCAG 1320  
 QY 1321 CTGACGCCACCTGACTCTGGCGCCGAAATCAGGTTACATCGGGCCGGAAGTCCGCGGAGAA 1380  
 Db 1321 CTGACGCCACCTGACTCTGGCGCCGAAATCAGGTTACATCGGGCCGGAAGTCCGCGGAGAA 1380  
 QY 1381 GATGCGATTTGGCAAGCAAGCTTCCGCAAGTACACCGACTACGCGAAGCTGTC 1440  
 Db 1381 GATGCGATTTGGCAAGCAAGCTTCCGCAAGTACACCGACTACGCGAAGCTGTC 1440  
 QY 1441 AAGCAAGAAATTCACAAAGTGGCCTGAGCATATGAGATGTTCTCCACCGCTTGGAG 1500  
 Db 1441 AAGCAAGAAATTCACAAAGTGGCCTGAGCATATGAGATGTTCTCCACCGCTTGGAG 1500  
 QY 1501 AGTGGCGCTACTTACGGGGTTCGATNTGCGGGCGCTGCTACAGTGGCGGATTCGCG 1560  
 Db 1501 AGTGGCGCTACTTACGGGGTTCGATNTGCGGGCGCTGCTACAGTGGCGGATTCGCG 1560  
 QY 1561 TTGGCCCCACCAAGAGTGGCAAGGCAACGAGCGCGAGCGGCTGGCGAAAGTCTCGAGC 1620  
 Db 1561 TTGGCCCCACCAAGAGTGGCAAGGCAACGAGCGCGAGCGGCTGGCGAAAGTCTCGAGC 1620  
 QY 1621 GTCTACGACAGATCTCTGCGCAACCGCGCTAGCATCGGAGAGTGTGCTTGGCGC 1680  
 Db 1621 GTCTACGACAGATCTCTGCGCAACCGCGCTAGCATCGGAGAGTGTGCTTGGCGC 1680  
 QY 1681 GGTACGCTAGCCATCGAAGGCGCGAAGCAGCAGAGTTCAGATGTGCGCGTCCCTTC 1740  
 Db 1681 GGTACGCTAGCCATCGAAGGCGCGAAGCAGCAGAGTTCAGATGTGCGCGTCCCTTC 1740  
 QY 1741 CTGAAGGCGCTGGCGATGCGACCGCGAGATGACCGAGAGTCTCTTCGACCGCTG 1800  
 Db 1741 CTGAAGGCGCTGGCGATGCGACCGCGAGATGACCGAGAGTCTCTTCGACCGCTG 1800  
 QY 1801 GAGCGCTGGCGATGGCTTCGCGACCTGGCAGAGAAAGAGTATGTTGGTGAAGCGGAA 1860  
 Db 1801 GAGCGCTGGCGATGGCTTCGCGACCTGGCAGAGAAAGAGTATGTTGGTGAAGCGGAA 1860  
 QY 1861 GAGATGCTGCTGGATCGTGGCAGCTGATGGGCTTAACCGCGCGGAAATGACGTGCTG 1920  
 Db 1861 GAGATGCTGCTGGATCGTGGCAGCTGATGGGCTTAACCGCGCGGAAATGACGTGCTG 1920  
 QY 1921 CTGGGGGATACCGCTATGGGCAACCACTATGTTGGTGGCCCAACACGCGGTATTCAC 1980  
 Db 1921 CTGGGGGATACCGCTATGGGCAACCACTATGTTGGTGGCCCAACACGCGGTATTCAC 1980  
 QY 1981 GATTTGAGGCGCATGTTGCAACAGCTTTTTGTGAACCTGACCATATGGGAGACAGC 2040  
 Db 1981 GATTTGAGGCGCATGTTGCAACAGCTTTTTGTGAACCTGACCATATGGGAGACAGC 2040  
 QY 2041 TGGAGGCGGTAGGTACCAAGCTGACCAACGCTTTTGTGAACCTGACCATATGGGAGAC 2100  
 Db 2041 TGGAGGCGGTAGGTACCAAGCTGACCAACGCTTTTGTGAACCTGACCATATGGGAGAC 2100  
 QY 2101 TGGAGCGGCTCGCGGGTGATCTGATTTGTTGGTTCGAACCTGCGCTACCTCTTACGCA 2160  
 Db 2101 TGGAGCGGCTCGCGGGTGATCTGATTTGTTGGTTCGAACCTGCGCTACCTCTTACGCA 2160  
 QY 2161 GAGTGTACGCCAGAGCATATAGCGGAGAGTTCGTGAGAGTCTGTCGCGCGCTGCG 2220  
 Db 2161 GAGTGTACGCCAGAGCATATAGCGGAGAGTTCGTGAGAGTCTGTCGCGCGCTGCG 2220  
 QY 2221 ACCAAGTGTAGAACCGCGACGCTTTCGACGCTGGCGTCTGATA 2262  
 Db 2221 ACCAAGTGTAGAACCGCGACGCTTTCGACGCTGGCGTCTGATA 2262

US-09-884-889-7  
 ; APPLICATION US/0984889  
 ; PATENT NO. US2002/0102680A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DIVERSA CORPORATION  
 ; APPLICANT: ROBERTSON, Dan  
 ; APPLICANT: SANYAL, Indrajit  
 ; APPLICANT: ADHIKARI, Robert  
 ; APPLICANT: SANYAL, Indrajit  
 ; FILE REFERENCE: DIV091100-005  
 ; CURRENT APPLICATION NUMBER: US/09/884, 889  
 ; PRIORITY FILING DATE: 2001-06-19  
 ; PRIORITY FILING DATE: US 09/412,347  
 ; PRIORITY FILING DATE: 1999-10-05  
 ; PRIORITY FILING DATE: US 08/951,844  
 ; PRIORITY FILING DATE: 1997-10-16  
 ; PRIORITY FILING DATE: US 08/674,887  
 ; NUMBER OF SEQ ID NOS: 8  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 7  
 ; LENGTH: 2238  
 ; TYPE: DNA  
 ; ORGANISM: Microscilla furvescens  
 ; US-09-884-889-7

Query Match 29.7%; Score 672.2; DB 10; Length 2238;

Best Local Similarity 59.5%; Pred. No. 1.3e-198;

Matches 1277; Conservative 0; Mismatches 808; Indels 60; Gaps 6;

Qy	150	CGGTACTTCCACAAAGATGTTGGCCGGAGAGGGTTGAACTCGATATTTGGCATCAGCA	209
Db	99	CGGCAACCAATAGGATTTGGCCGACACACCTCAACTTGGGCTTTACGCGACCA	158
Qy	210	AGATCGCAATACAGACCCGATGGATTCGCGATTTCAACTACCGTAGAGAGTAGCAGCT	269
Db	159	TTCAATGCTGTTTGGGACCAACAGACCCGATTTGCACTATGCGGAGAGCTTTAAGAGCT	218
Qy	270	CGATTTGACGCGCTGAGAAAGATGTCACGCGTTGATGACGATGACGACGAGTGGTG	329
Db	219	AGATCTGGACGGGTTTAAAGAGGCTGGCGAGCCCTANTGACAGATTCACAGACTGGTG	278
Qy	330	GCCCGCTGACTTGGGGGCACTACGGCGGTTTGATGATCGGTATGGCTTGGCACTCCGCTGG	389
Db	279	GCGACGATATGCTATGATGCGCGGCTTCCTTTATACGCTGGCGGACAGCGCGG	338
Qy	390	CAGCTACGATATGCTATGATGCGCGGCTTCCTTTATACGCTGGCGGACAGCGCGG	449
Db	339	CACCTACGCTATGCTGATGCGCGGCTTCCTTTATACGCTGGCGGACAGCGCGG	398
Qy	450	GCTCAACTCTGCGCGGACAGCTGACGCTGGATTAAGCGCGCTGCTGCTGGCGGAT	509
Db	399	TCTCAATAGCTGGCGGACAGCTGCAATCTGGATTAAGCGCGCTGCTGCTGGCGGAT	458
Qy	510	CAGACGATGAGGACGCAAAATGCTGGCGGACAGCTGCAATCTGGATTAAGCGCGCT	569
Db	459	CAAAACAAATACGCTGGAATAATCTCTGGCGGATTAATGATCTACACGGAAGCT	518
Qy	570	GCTTATGATGCTTACGCTTACGCTTCTCTTTCGGCGGCGGTGATATTTG	629
Db	519	AGCTCTGGAATATGCGCTTAAATCTTTGCTGGAGTGGCAGACGATGATG	578
Qy	630	GGAGGCTGAGAGATGATATCTGGGAGCGAGAACCGATGCTTGGGA	632
Db	579	GGAGGCTGAGAGATGATATCTGGGAGCGAGAACCGATGCTTGGGA	632
Qy	690	AGGCTACGCGGATGACAAAGCGGAGACGATGGAACCCGCTGGCGGCTGTCAAAAT	749
Db	633	GCGCTATGAGAGTGGACCGAGAGC-----TCGAAATCCCTTGGGAGCGGTCAAAAT	683
Qy	750	GGGCTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	809
Db	684	GGGCTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	743

Qy	810	ACAGCAGCTACTTGAACCTTCCGCCCTATGCGGATGAGCAACAGCAAAAGAACCCGACCT	869
Db	744	GGTGATATTCGTGAGACTTTTGGCGGANTGGCAATGATGACGAGAAACCCGTGCTCT	803
Qy	870	CACAGCTGGCGGCACACCGTCCGTAATTTGTCACGGTAATGCAATGCTCTGCTCT--T	926
Db	804	CATAGCGGTGGACACACCTTCGGAATAACCCATGCTGCTGCCGATGCGGAGAAATATCT	863
Qy	927	AGCCCTCTGACCAAAAGCCTCTGAGTTGAAACACAGGCGGTAGGTGGCGCAACCGTA	986
Db	864	GGSCCGAGAGCCTCGCCGCGAGGTATTTGAAGAAATGAGCTGGGGTGGAAACACCTA	923
Qy	987	CATGCAAGGCAAGCAAGCAACGCCGTGACTCGGGTATGCAAGGTGCTTGGACACCA	1046
Db	924	CGGCACCCGACAGCTGGCGGATACCAATCAACAGTGGACTAGAGGGCGCTTGGACCAAG	983
Qy	1047	CGCCAGCAAAATTCGATATGCGGCTATTTGACCTGCTGTCGCTGACATTTGGGAATGAA	1106
Db	984	CCCTACTCAATGGAGCAATAACTTTTTTGAACACCTCTTTGGTTACGAGTGGGACTTAC	1043
Qy	1107	AAGAGTCTGCGGTGCCCGCATTTGGGAACCGATGCAATCAAAAGAGAAACAAAGCC	1166
Db	1044	CAAAAGTCCAGCTTGGAGCTTATCAGTGGAAACCAAAAGAGCGTGGCGGGCTGGCACAT	1103
Qy	1167	GTGGACGACGAGCGCCCTATTTTCGCGACAGACGATGATGAGAGTGGGATGAA	1226
Db	1104	ACCGGATGCAATGATCCGACGCTATGCGGACGCTGCTCCATTTATGCTACATCGGACCTGG	1163
Qy	1227	GATAAAGTAAATTCGAGACTATCGGGCTATCTCGGAAATATTCATGCGGATCCTGAGTA	1286
Db	1164	GTGCGCATGACCTCTGATGAGAAATATTTTCGACGCGTACTATGAAACCTGATGA	1223
Qy	1287	CTTCAGAAAGATTTTCGCGAGCGGTGTCAGCTGACGACGCTGACCTGGCGCGGAA	1346
Db	1224	GTTTGCAGATGCTTTTCGGAAGACATGTCACAAATGACACACAGATATGGGACCAA	1283
Qy	1347	ATCAGCTTACATGCGCCCGGAAAGCTGGCGAGAGACCTGATTTGGCAAGACCGCATTC	1406
Db	1284	GGTGGCTACCTGGCGGACGAGTGGCTCTCAGGAGACCTCTCTGGCAGACCTTATCC	1343
Qy	1407	GGCAGGTAAAC-----ACGCACTACTGCGAAGATGTCACGACGAGAAATTCG	1454
Db	1344	AGATGTAGCCATCTCTTGTAGAGCAAAACGATATTTGAAGCCTTAAAGACAAAATCT	1403
Qy	1455	ACAAAGTGGCTGACATTAAGTGAAGTGTCTCCGCGCTGGGACAGTGGCGGACTCTA	1514
Db	1404	GGATTCGCGATCGGTGACGAGTGTGACGAGCGATGGCTTTCGCTCTCTCTTT	1463
Qy	1515	TGCGGTTTCGATATGCGCGCGGTGCTTACGCTGCGCGGATTCGCTTGGCGCCACAGAA	1574
Db	1464	TAGAAACTCTGCAAGAGCGCGGTGCGCAACGCTGCAAGTATACGACTGGCGCCCAAAA	1523
Qy	1575	CGACTGGCAAGGCAACGACCGCGAGCGCTGGCGGAAGTGTGAG-----	1619
Db	1524	AGCTTGGAGGACCAACCTCTCCAGGACTTCCGAGCTACTCGAGCTACGAGACTTAAGAGTAT	1583
Qy	1620	-----GCTCTACGACGAGATCTCTCGGACACCGCGCTAGCATGCGGAGCTGAT	1670
Db	1584	CCAGSAGAGACTTTTAAACCAGGGCGAATCAGATTAACAAAGCAGTATCTGTTGGCGACCTGAT	1643
Qy	1671	CTTCTGGCGGTAGCGTAGGATAGCAATGAGAAAGCGCGGAAAGCAGGATTCAGGATGGC	1730
Db	1644	TGCTTGGCGGTGCTGGGCTGTAGNAAAGCTTCNAAATATGCTGGCATGGGTGCA	1703
Qy	1731	CGTTCCTTCTGAAAGCGGTGGGATGCGGACCGCGGAGATGACGACGAGCTCTCT	1790
Db	1704	GGTGGCTTTCACCGCGGACGAGCGGCTGCAACGCTGAGCAACCGATGTGGAGCTTT	1763
Qy	1791	CSCACCGCTGGAGCGCTGGCGGATGGCTTCCGCACTTCCGAGAGAAAGATGATGGT	1850
Db	1764	CGAGCACTAGGCGCGCTGACCGCTTTGAAACTACTACTTAAACCGAGCTAAGCT	1823



```

Oy 1528 ATGCGGCGGCGCTTAACGCTGCGCCGATTCGCTTGGCCCTCCACAGAACAGCTGGCAGGCG 1587
Dy 39037 AAGCGCGGCGGCGCCACGCTGTGTGTCGTCGTCGCTCGACGACAGTGGCTGGGAGTGC 38978
Oy 1588 AACGAGCGGGA--CGCCCTGGGGAAGTCTGCTGAGGCTCTACGACGATCT-----1636
Dy 38977 ACGACCCCGGAGGATCTGGCGAAGTGTATCGCACCTCTGGAGAGATCCAGGAGTA 38918
Oy 1637 -----CTCCGCGCGGCTAGGCTGCGGAGCTGTATCTGTGGCG 1680
Dy 38917 TTCAACTCGCGGCGCGCGGGAATCAAAAGTGTCTCGCGACCTCGCTGCTCGGT 38958
Oy 1681 GGTAGCTAGCATCGAAGAACCGCGGAAGCAGAGTGTAGCTGTGGGCTTCCTCTC 1740
Dy 38857 GCTGTGCGCCGTCATAGAGAACGACAAAGGGGCTGGCCACACATCAACGCTGCGCTC 38798
Oy 1741 CTGAAGCGGCTGGCGCTGCGCGGAGATGACGACGACGACGCTCTGCGACCGCTG 1800
Dy 38757 ACCCGCGCGCGCACGCTGCTCGCAGCAACAAACGACGTGGAATCTCTTGCGCTGCTG 38738
Oy 1801 GAGCGCTGCGCGATGCTCTTCGCACTGCGCAGGAAGAAGATGTGTGTGAAGCGCGAA 1860
Dy 38737 GAGCCAGGCGAGATGCTCTCGACATCTCTCGGAAGAGCGCAACGCTGTGCGCGCGAG 38678
Oy 1861 GAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
Dy 38677 TACATCTGCTCTGACACAGGACCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 38618
Oy 1921 CTGGCGGCTATGCGGCTACTGGGCGACCACTATGCTGGCACCAACACGCGGCTATTCAAC 1980
Dy 38617 GTAGGTGGCTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 38558
Oy 1981 GATTGTGAAGGCCATGACACAGGACTTTTGTGTAACCTGACCTGACGATGAGGACAGC 2040
Dy 38557 GAGCGCTCCGAGTCACTGACACAGGACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 38498
Oy 2041 TGAAGC---CGTAGGTAGCAAGCTCTAGCAATCCCGCACCAACACGCGGCTGCGGTG 2097
Dy 38497 TGGAGGCTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 38438
Oy 2098 AAGTGAAGCGCTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2157
Dy 38437 AAGTGAAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 38378
Oy 2158 GCGAGATGTACGCCAGAGAGATACAGCGGGAAGTGTGTCAGAGATCTGCTGCGCGCG 2217
Dy 38377 GTCAGGCTGTATGGCGCTGTGTCAGCGGACCGCGAGTCTGTGCGAGACTTGTGCTGCTG 38318
Oy 2218 TGGACAAAGTATGACACGCGACCTTTTCGAGCT 2252
Dy 38317 TGGACAGCTGTATGACCTCGACAGTGTGCGAGCT 38283

RESULT 4
US-09-940-925A-144
: Sequence 144, Application US/09940925A
: Publication No. US20030054338A1
: GENERAL INFORMATION:
: APPLICANT: BROW, MARY ANN D.
: INVENTOR: LYAMICHEV, VICTOR I.
: TITLE OF INVENTION: DETECTION AND IDENTIFICATION OF
: PATHOGENS
: NUMBER OF SEQUENCES: 165
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
FIRMWARE: PC-DOS/MS-DOS
SOFTWARE: PLOT RELEASE #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940, 925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER S.
REFERENCE/DOCKET NUMBER: FOXS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 144:
SEQUENCE CHARACTERISTICS:
TYPE: Nucleic Acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 144:
US-09-940-925A-144
Query Match
Best Local Similarity 54.3%; Pred. No. 4, 7e-32;
Matches 330; Conservative 0; Mismatches 263; Indels 15; Gaps 2;
Oy 994 GCGAAGCGACGACGCGCTGACCTGGGATTCGGAAGTGTCTTGACCAACACCCACG 1053
Dy 16 GCGACGCTAAGAGCGCGGATCACCCGCGCTCGGCTGTATGGAGGACGCCGACG 75
Oy 1054 AATTCGATATGGCTATTTCGACCTGCTTGGCTACATTTGGGACTGAAAGAGT 1113
Dy 76 AATTCGACACAGATTTCTCGAGATCTCTACGCTACGAGTGGAGCTGACGAGAGC 135
Oy 1114 CTTCCGCTGCGCCACCATTTGGGACCGGATTTGACATCAAAAAGAAAGCCGCTTGAC 1173
Dy 136 CTTCCGCTGCGCTTGGCTATTCACCGCCAGGAGCGCGCTGCGGACATCCCGGAGC 195
Oy 1174 GCGAGCGACCTCTTATGCGCACACCGCATCATCGGATGCGGATGCGGATGTCG 1233
Dy 196 CCGTTTGGCGGGCC---AGCGCGCTCCCGAGCATGCGCACATCTCTGCTGCGG 252
Oy 1234 GTAAATCGACCTATCGCTCTATCTCGGAAAATTCATGCGCGATCTCGAGTCTTCAAG 1293
Dy 253 GTGATCTGATCTTATGAGCGGATACGCGCTGCTGCTGCGACCCCGGAAATTTGCC 312
Oy 1294 AAACCTTTGCGAAGGGGTGTTCGACGTCGACGACCGTGAACCTGGGCGCGAATCACT 1353
Dy 313 GACGAGTTTCCCAAGGGCTGTGTACAGCTGATCCACCGAGACATGGTCCGTTGCGA 372
Oy 1354 TACATGCGCGGAAGTGGCGGCGAGAGCATGATTTGGCAAGACCGGATTCGCGAGGT 1413
Dy 373 TACTTGGGCGCTGTCTCCACAGACAGACCTGCTGTGGGAGATCCGCTGCGGTC 432
Oy 1414 AACACGCTACTGCGGAAGAGTGG-----TCGAGCAGAAATTCGACNAAGT 491
Dy 433 AGCCACGACCTCTGCGCGAGCGGAGTTTGAAGAGCAGAGATCCGCGCATCG 492
Oy 1462 GCGCTGAGCATTAGTGAGATGTCTCCACCGCTTGGGACAGTGCCTGCTACTTATCCGGT 1521
Dy 493 GATTTGATCTGTCTACAGTATTGTTCAGCGCATGGGCGCGCTGCTGCTCGGTGT 552
Oy 1522 TCCGATATGCGCGGGGTGCTTACGCTGCCCGCATTCGTTGGCCCGCACAGACGATGG 1581
Dy 553 AGCCACAGCGCGCGGCGCCACAGCGTGTGCTCATCGCTGCAACCCACAAGTCCGGTGG 612
Oy 1582 CAGGCGAA 1589
Dy 613 GAGGTCAA 620

```

```

RESULT 5
US-09-940-925A-146
: Sequence 146, Application US/09940925A
: Publication No. US20030054338A1
: GENERAL INFORMATION:
: APPLICANT: BROW, MARY ANN D.
: LYMICHEV, VICTOR I.
:
: TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
: PATHOGENS
:
: NUMBER OF SEQUENCES: 165
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MELEN & CARROLL
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/940,925A
: FILING DATE: 10-Jun-2002
: CLASSIFICATION: <unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: CARROLL, PETER G.
: REGISTRATION NUMBER: 32,837
: REFERENCE/DOCKET NUMBER: FORS-01756
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 397-8338
: TELEFAX: (415) 397-8338
:
: INFORMATION FOR SEQ ID NO: 146:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 620 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 146:
US-09-940-925A-146

Query Match 6.1%; Score 137.2; DB 11; Length 620;
Best Local Similarity 54.3%; Pred. No. 4.7e-32;
Matches 330; Conservative 0; Mismatches 263; Indels 15; Gaps 2;

QY 994 GGCAGGCAAGCAACCGCGTGAACCTCTGGATATGCAAGAGTGTCTGGACCAACCCACG 1053
DB 16 GGAAACCGGTAAAGACCGATCATCCAGCGCATCGAGGTCTGTATGGACCAACCCGACG 75

QY 1054 AAATTCGATAGGCGTATTTGGACCTCGTGTTCGGCTACATTCGGAACTGAAAAGAGT 1113
DB 76 AATTCGGACACGATCTTCGTGGATCTCTGTACGCTGACGTGGAGCTGACGAGAGC 135

QY 1114 CTTGGTCCGACCCACCATCTGGACATCGATGATCAAAAGAGAAACCAAGCGGTGAC 1173
DB 136 CTTGCTCGCCCTTTGGCAATACACCCGCAAGGAGCGCGCGGTGCGCGGACCATCCGAG 195

QY 1174 GCAGGAGCCCTCTATTCGCAACCAACCGCATCATGACGATGCGGATATGGCGATTAAG 1233
DB 196 CGTTGCGCGGCC---AGGCGCTCCCGAGCATGTGCGCTGACTCTGCTGCGCG 252

QY 1234 GTAATTCGACCATATGCGGCTATCTCGGAAATTCATCGGATCTCTGACTACTCAAG 1293
DB 253 GTGATTCGCAATCATGACGCGGATCATCCGCTCGCTGGACACCCCGAGGAATTGCC 312

QY 1294 AAMACTTTCCGCAAGCGCTGTTCACTGACGTGACGACGATGCTGACGTGGGCGGAAATCACT 1353
DB 313 GACGATTTGCCCAAGGCTGTTGACAGCTGATCCAGCGAGCATATGGTCCGTGGGAGA 372

```



[illegible]

```

: TELEPHONE: (415) 705-8410
: TRUFPAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 150:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 620 base pairs
:   TYPE: nucleic acid
:   STRAIN: MESS-2 double
:   TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 150:
US-09-940-925A-150

Query Match          6.1%; Score 137.2; DB 11; Length 620;
Best Local Similarity 54.3%; Pred. No. 4.7e-32;
Matches 330; Conservative 0; Mismatches 26; Indels 15; Gaps 2;

QY 994 GGC+AAGCAGCAACGACGGTGAACCTCGGGGTATCGAAGTGTCTGTGACCAACCAACCCGACG 1053
Db      |||||
QY 605 GGAACCCGGTAAAGAGCGCGATACCAACCGGCATCGAGGTCTGTATGACGACACACCCCGACG 546
Db      |||||
QY 1054 AAMTTCGNATGGGCTATTTCGACCTGCTTGGCTTCAGATTGGGAATCTGMAAAGAT 1113
Db      |||||
QY 545 AATCCGACACAGATTTCTCGCATCTCTGTAGGCTGTGGGCTGTGAGGCTGTGACGAGAC 486
Db      |||||
QY 1114 CATTGCGGTCGCCACCATTCGGGAACCGATTGACATCAAAAAGGAAACAAAGCCGGTTGAC 1173
Db      |||||
QY 485 CTGTGCTGGGCGTTTGGCAATACACCGCAGACAGGCGCGGTGCCGGCACCATCTCCGGAC 426
Db      |||||
QY 1174 GCCAGCGACCCCTCTATTTCGCCACACCCGATCATGACCGCATGCGGATATGGCATTAAG 1233
Db      |||||
QY 425 CTTTGGCGGGCG---AGCGGCTCTCCCGACGATGTGGCCATCTGACCTCTCCGCG 369
Db      |||||
QY 1234 GTAAATCGGACTATCGGCTATCTCGGAAATATCATGCGCGATCTCTGTACTACTCAAG 1293
Db      |||||
QY 368 GTGGATCCGATCTATGAGCGGATACCGGCTGCGTGGCTGGAACACCCCGAGGAATTTGGCC 309
Db      |||||
QY 1294 AAGACTTTGCGAAGGCGGTTCAGCTGACGACCGTACATCTGGGCCGACGATCTGGGCCGGAACATCACT 1353
Db      |||||
QY 308 AGGACGACGCTGTGTCAGAGCTGATCAACGATGATCAACGAGTACATGCGTGTGGGAG 249
Db      |||||
QY 1354 TACATCTGGCGCGGCTGATGCCGGCAGAGACCTGATTTTGGCAAGACGCGGATTCGGCGAGT 1413
Db      |||||
QY 248 TACCTTTGGCGCGCTGTTCCCAAGCAGACCTCTCTGTGGCAGGATTCGGGTCTCTCGGCT 189
Db      |||||
QY 1414 AACACGCACTACTCGAAGAGATGG-----TGAACGAAAMTTGCACAMGTT 1461
Db      |||||
QY 188 AGCCACAGCTGCTGGCGAACCAGATATGCCAGCTTATAGAGCGAATCTTGGCATCG 129
Db      |||||
QY 1462 GGCCTTGACATATAGTAGATGTCTCCACGCTTGGGACAGTGGCCGCTACTTATCGGGT 1521
Db      |||||
QY 128 GGAAATGACTGTCTACAGCTTAGTTTCGACCGCATGGGCGGGCGCTGCTGCTCGGTGT 69
Db      |||||
QY 1522 TCGGATTTGGCGGCGGTGTACGAGTCCGCGATTTGGCTTTGGCCCCCAGACGACGATGG 1581
Db      |||||
QY 68 AGCGACAGCGGGCGGCGCCACAGTGGTGGCATCTCCGCTACGCCCTACAGTCGGGTGG 9
Db      |||||
QY 1582 CAGGCGCAA 1589
Db      |||||
QY 8 GAGGTCAA 1

RESULT 8
US-09-941-134A-144
: Sequence 144, Application US/09941193A
: Publication No. US20030108873A1
: GENERAL INFORMATION:
: APPLICANT: BROWN, MARY ANN D.
: LYAMICHEV, YURI I.
: OLIVE, DAVID M.
: TITLE OF INVENTION: DETECTION AND IDENTIFICATION OF
: BACTERIAL PATHOGENS
: NUMBER OF SEQUENCES: 165
: CORRESPONDENCE ADDRESS:

```



```

Db 253 GTGGCTCGATCTATGAGCGATCAAGCGCTGCTGCTGAGACACCCCGAGGATGGGC 312
Qy 1294 AAAACTTTCCGAGAGCGGTGTTCAAGCTGAGCGACCGCTGAGCTGGGCGCCGAATAACGT 1353
Db 313 GACGAGTTCCGCAAGCGCTGGTACAAAGCTGATCCACGAGACATGGTCCGTGCGAGA 372
Qy 1354 FACATCGCGCGAGAGCGGCGGACAGACCTGATTGGCAGACCGGATTCGGCGAGGT 1413
Db 373 TACCTTGGGCGCGTGGTCCCAAGCAGACCGCTGCTGGCGATCCGCTCCCTGCGGTC 432
Qy 1414 AACACCGCACTACTGCGAAGAATGGG-----TCAAGCAGAAATATGCAAAAGT 1461
Db 433 AGCCACGACCTGCTGGCGAGCGGACGAGATTCGACGCTTAAAGCGCAGATCTCGCATCG 492
Qy 1462 GCGCTCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1521
Db 493 GGATGACTGTCTCACAGTAGTTTGCAGCGCATGGGCGCGCGTGGTGGTGGTGGTGGT 552
Qy 1522 TCGCATGAGCGCGCGTGTCTAACGGTGGCCGCAATTCGCTTGGCGCCACAGAACAGTGG 1581
Db 553 AGCCACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 612
Qy 1582 CAGGCGCAA 1589
Db 613 GAGTCAA 620

RESULT 10
US-09-941-193A-148/c
: Sequence 148, Application US/09/941/193A
: Publication No. US20030108873A1
: GENERAL INFORMATION:
: APPLICANT: BROW, MARY ANN D.
: OLIVE, DAVID M.
: TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
: NUMBER OF SEQUENCES: 165
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/941/193A
: FILING DATE: 28-Aug-2001
: PRIORITY DATE: 28-Aug-2001
: ATTORNEY: PETER G. CARROLL
: NAME: CARROLL, PETER G.
: REGISTRATION NUMBER: 32,837
: REFERENCE/DOCKET NUMBER: FORS-01756
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 148:
: SEQUENCE LENGTH: 620 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 148:
US-09-941-193A-148
Query Match 6.1%; Score 137.2; DB 11; Length 620;

```

```

Beat Local Similarity 54.3%; Pred. No. 4.7e-32;
Matches 330; Conservative 0; Mismatches 26; Indels 15; Gaps 2;
Qy 994 GCGAAGCGACGACGCTGAGCTGCGGTATCCGAAGTGTCTTGACACCAACCCGACG 1053
Db 605 GGAACCGGTGAAGCAGCGGATCACCCAGCTCGAGGTGCTGTATGGACGAACACCCGACG 546
Qy 1054 AAATTCGATATGCGCTATTTCGACCTGCTGTCGGCTACATTTGGAATCGAAAAGAGT 1113
Db 1054 AATTCGACCAAGTTCCTCGAGATCTCTGAGCTGAGTGGAGAGGCTGCGAGAGC 486
Qy 1114 CTTCCGCGTCCGACCATTTGSGACGATGACATCAAAAGAAAGAAACGCGGTGAC 1173
Db 485 CTTCTCGCGCTTGGCAATAACACCCGACGACGAGCGGCGGTGCGGCGACCATCCCGAC 426
Qy 1174 GCGAGCGACCCCTATTCGCGCAACCGCATCATGACCGATGCGGATGGCGATAAAG 1233
Db 425 CGGTTCGGCGGCGC---AGGCGCTCCCGAGATGCTGCGCATGACTCTCTGCTGCGG 369
Qy 1234 GTAAATTCGACCTTATCGGCTATCTCGGCAAAATTCATGCGCATCTCTGAGTACTTCAAG 1293
Db 368 GTGATCGCATCTATGAGCGGATCAGCGCTGCTGCGTGGACACCCCGAGGAATTCGCC 309
Qy 1294 AAACCTTCGCGAAGCGGTGTTCAAGCTGACGACCGCTGACCTGGCGCCGGAATCAGT 1353
Db 308 GACGAGTTCGCGAAGCGGTGTTCAAGCTGACGACCGCTGCTGCGTGGACACCGCATTCG 1413
Qy 1354 TACATCGCGCGGAAAGTGGCGAGACGACCTGATTTGGCAGACCGGATTCGCGCAGGT 249
Db 248 TACCTTGGGCGCGTGGTCCCAAGCAGACCGCTGCTGCGCAGGATTCGCGTCTCGGTC 189
Qy 1414 ACACGAGTACTGCGGAAGAATGG-----TCAAGCAGAAATTCGCAAAAGT 1461
Db 188 AGCCACGACCTGCTCGCGAGCGCGAGATGCGACGCTTAAAGCGCAAGATCCGGGCGTCG 129
Qy 1462 GCGCTGAGCATTAAGTAGATGTTCTCCACCGTTGGGACAGTGGCGGTATTCGCGGT 1521
Db 128 GATTAAGTGTCTCTCACAGAGTAGTTTCGACCGCATGGGCGCGCGCTGCTGCTGCGGT 69
Qy 1522 TCGCATGAGCGCGGCGGTGCTAACGGTCCCGCATTCGCTTGGCGCCACAGACGAGTGG 1581
Db 68 AGCCACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9
Qy 1582 CAGGCGCAA 1589
Db 8 GAGTCAA 1

RESULT 11
US-09-941-193A-150/c
: Sequence 150, Application US/09/941/193A
: Publication No. US20030108873A1
: GENERAL INFORMATION:
: APPLICANT: BROW, MARY ANN D.
: LYAMICHEV, VICTOR I.
: OLIVE, DAVID M.
: TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
: NUMBER OF SEQUENCES: 165
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/941/193A
: FILING DATE: 28-Aug-2001
: PRIORITY DATE: 28-Aug-2001
: ATTORNEY: PETER G. CARROLL
: NAME: CARROLL, PETER G.
: REGISTRATION NUMBER: 32,837
: REFERENCE/DOCKET NUMBER: FORS-01756
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 149:
: SEQUENCE LENGTH: 620 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 149:
US-09-941-193A-149
Query Match 6.1%; Score 137.2; DB 11; Length 620;

```

FILING DATE: 26-Aug-2001  
 CLASSIFICATION: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: CARROLL, PETER G.  
 REGISTRATION NUMBER: 32,837  
 REFERENCE/DOCKET NUMBER: FOS-01756  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION SOURCE:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 620 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 SEQUENCE DESCRIPTION: SEQ ID NO: 150:  
 US-09-941-193A-150

Query Match 6.1%; Score 137.2; DB 11; Length 620;  
 Best Local Similarity 54.3%; Pred. No. 4,7e-32;  
 Matches 330; Conservative 0; Mismatches 263; Indels 15; Gaps 2;

Qy	994	GCACAGCAGCAACCGCGTGCACCTCGGCTATCGAAGAGTCTTGGACCAACCAACCCCGC	1053
Db	605	GGACCGGTATGACGACATCGCCAGCGATCGAGTGTGATGACGACACCCCGCG	546
Qy	1054	AAATTCGATAGGCTATTTCGACCTCGCTTGCGCTACACANTTGGGAAGTGAAGAGCT	1113
Db	945	AAATGGGACAGCACTTTCCTCGAGATCTGTACGCTACGATGGAGCTGACGAGAGC	486
Qy	1114	CTGCGGCGTCCACACCTTGGACCGCTTGCATCTAAMAGGAAACAGCGGGTGTAC	1173
Db	485	CTGCTTGGCGCTTGGCAATACACGCCAAGGAGCGCGCGTGGCGCAACCATCCCGAC	426
Qy	1174	GCACAGCAGCCCTCTATTTCGCCACACACCGCATGATGCGATGGGATGGCGATMAAG	1233
Db	425	CCGTGTCGGCGGGCC---AGGGCGCTTCCCGGAGTGTGCGCATCGACTCTCGTCGG	369
Qy	1234	GTAATTCGACACCTATCGCTATCTGCGAAMATTCATGCGCATCTGAGTACTCAG	1293
Db	368	GTGGATCTGATGAGCGGATACACGCTGCTGCTGTGACATCCCGAGGANTTGGCC	309
Qy	1294	AAATCTTTCGCGAGCTGGTTCAGCTGACGACCTGTGACTCTGGCGCGGAAATCACT	1353
Db	308	GACGAGTGTGCCACAGCGCTGTGACAGCTGATCAGCTACGACAGACATGGGTCCGTG	249
Qy	1354	TACATGCGCCCGAGATCGCGGAGAGACCTGTTTGGCGAGACCGATTCGCGAGGT	1413
Db	248	TACCTTGGCGCGCTGGTCCCGAAGCAGACCTGCTGTGGAGAGATCGGCTCGCTGG	189
Qy	1414	AACACCGACTACTCCGAGAAAGTGG---TCACGAGAAATTCGACAAAGT	1461
Db	188	AGCCAGACCTCTGCTGGGAGAGCGGAGATTCGCCAGCTCTAAGACCATCTGGCATG	129
Qy	1462	GGCTCGAGCATCTGATGAGTGGCTCTACCGCTGACAGTCCGCTCTACTTCCGGGT	1521
Db	128	GGATTTGACTCTCACAAGTATTTCACCGCATGGCGGCGCGCTGCTGCTTCCGTGGT	69
Qy	1522	TCGAGATGCGCGCGGGTGTACAGGTGGCGGATATGGTGTGGCCCGACAGAAAGATGG	1581
Db	68	AGCAGACAGCGCGCGCGCGCGACAGCGTGGTGGCTGCTGCGCTGACGACCAAGT	9
Qy	1582	CAGGCGCA 1589	
Db	8	GAGGTCAA 1	

RESULT 12  
 US-09-940-925A-143  
 Subject: 143 Application US/09940925A  
 Publication No. US2003005438A1  
 GENERAL INFORMATION:

APPLICANT: BROW, MARY ANN D.  
 OLIVE, DAVID M. DETECTION AND IDENTIFICATION OF  
 TITLE OF INVENTION: PATHOGENS  
 NUMBER OF SEQUENCES: 165  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MEDLEN & CARROLL  
 STREET: 220 MONTGOMERY STREET, SUITE 2200  
 CITY: SAN FRANCISCO  
 STATE: CALIFORNIA  
 COUNTRY: UNITED STATES OF AMERICA  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA: US/09/940,925A  
 FILING DATE: 10-JUN-2002  
 CLASSIFICATION: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: CARROLL, PETER G.  
 REGISTRATION NUMBER: 32,837  
 REFERENCE/DOCKET NUMBER: FOS-01756  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 143:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 620 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 SEQUENCE DESCRIPTION: SEQ ID NO: 143:  
 US-09-940-925A-143

Query Match 5.0%; Score 135.6; DB 11; Length 620;  
 Best Local Similarity 54.1%; Pred. No. 1.5e-31;  
 Matches 329; Conservative 0; Mismatches 284; Indels 15; Gaps 2;

Qy	994	GCACAGCAGCAACCGCGTGCACCTCGGCTATCGAAGTCTTGGACCAACCAACCCCGC	1053
Db	16	GGAAACCGGTAAAGACCGCATACACGAGCGCGCATCGAGTCTGATGGAACCAACCCCGAG	75
Qy	1054	AAATTCGATAGGCTATTTCGACCTCGCTTGGCTACANTTGGGAACTGAAAGAGT	1113
Db	76	AANTGGACACACTTTCCTGAGATCTGTAGGCTATGAGTATGATCAAAAGAAACAGCGGTGAC	135
Qy	1114	CTGCGCGCTGCCACCATCTGGACCGATGATGATCAAAAGGAAACAGCGGTGAC	1173
Db	136	CTGTCTGCGCTTGGCAATACACGCCAAGGAGCGCGCGTGGCGGACCATCCCGGAC	195
Qy	1174	GCACAGCAGCCCTCTATTTCGCCACACCCCGATGACCGATGCGGATGCGCATMAAG	1233
Db	196	CGGTCGGCGGGCC---AGGGCGCTTCCCGGAGTGTGCGCATCTGCGCTGCTGGTGG	252
Qy	1234	GTAATTCGACACCTATCGGCTCTATTCGCGAAMATTCATGCGCATCTGAGTACTCAG	1293
Db	253	GTGGATCGCATCTATGAGCGGATTCACCGCTGCTGCTGGAACACCCCGAGGANTTGGCC	312
Qy	1294	AAATCTTTCGCGAGCTGGTGTGACAGCTGATGACGCGTGGCTGCGGCGGAAATCACT	1353
Db	313	GACGAGTTGCCAAGCGCTGCTACACCTGATCCACGAGACATGGGTCCGTTCGAGA	372
Qy	1354	TACATGCGCCCGGAGTCCCGGACAGACCTGATTTGCAAGACCGCATTCCTGGCAGGT	1413
Db	373	TACCTTGGCGCGCTGGTCCCGAAGACAGACCTCTGTGGAGATCCGGTCCCTCGCGTGC	432
Qy	1414	AACACCGACTACTCCGAGAAAGTGG-----TCACGAGAAATTCGACAAAGT	1461





Search completed: October 8, 2003, 19:55:23  
Job time : 586.569 secs

**THIS PAGE BLANK (USPTO)**



GenCores version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 07:49:13 ; Search time 4688.37 Seconds  
(without alignments)  
11720.187 Million cell updates/sec

Title: US-09-884-889-5

Perfect score: 2262  
Sequence: 1 atgaataacgcacgcgtga.....gtttcgcgcgcgcgttaa 2262

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:\*

1: em.estha:\*  
2: em.esthum:\*  
3: 9b.est1:\*  
4: em.estnu:\*  
5: em.estvot:\*  
6: em.estpl:\*  
7: em.estro:\*  
8: em.htc:\*  
9: 9b.est1:\*  
10: 9b.htc:\*  
11: 9b.htc:\*  
12: 9b.est3:\*  
13: 9b.est4:\*  
14: 9b.est5:\*  
15: em.estom:\*  
16: em.estom:\*  
17: em.gss.hum:\*  
18: em.gss.hum:\*  
19: em.gss.dlu:\*  
20: em.gss.vrt:\*  
21: em.gss.fun:\*  
22: em.gss.mam:\*  
23: em.gss.mus:\*  
24: em.gss.pro:\*  
25: em.gss.pod:\*  
26: em.gss.pod:\*  
27: em.gss.vrt:\*  
28: 9b.gsl1:\*  
29: 9b.gsl2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	285.2	12.6	1102	28 AF075865	AF075865
2	261.4	11.6	707	13 BQ751801	BQ751801
3	261.4	11.6	707	13 BQ751801	BQ751801
4	257.8	11.4	562	10 BQ370112	BQ370112

#### SUMMARIES

#### ALIGNMENTS

#### RESULT 1

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

1102 bp DNA linear GSS 29-MUG-2000  
AF075865 Salmonella typhimurium LT2 Lambda DASH II Salmonella  
typhimurium genomic clone 1079-77, genomic survey sequence.  
AF075865  
GSS.  
Salmonella typhimurium  
Salmonella typhimurium  
Salmonella typhimurium  
Enterobacteriaceae: Salmonella.  
REFERENCE  
1 (bases 1 to 1102)  
Wong, R.M.-Y., Wong, K.K., Benson, N.R. and McClelland, M.  
Sample sequencing of a Salmonella typhimurium LT2 lambda library:  
comparison to the Escherichia coli K12 genome  
JOURNAL  
FEMS Microbiol. Lett. 173 (2), 411-423 (1999)  
MEDLINE  
92433757  
10727710  
COMMENT  
McClelland M  
Molecular Biology  
Sidney Kimmel Cancer Center  
3099 Science Park Road, San Diego, CA 92121, USA  
Email: mclelland@fscf.sdsu.edu  
Class: shotgun.  
Location/Qualifiers  
1..1102  
source

#### ALIGNMENTS

#### RESULT 1

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

linear

GSS 29-MUG-2000

AF075865

1102 bp

DNA

Query Match	11.6%	Score 261.4	DB 13	Length 707
Best Local Similarity	68.5%	Pred. No. 2.4e-60		
Matches 361	Conservative	0	Mismatches 166	Indels 0
Gaps	0			





BASE COUNT	131 a	222 c	176 g	110 t	recombinants. *
ORIGIN					
Query Match		10.4%	Score 236;	DB 13;	Length 639;
Best Local Similarity		68.5%;	Pred. No. 2.1e-53;		
Matches 326;		Conservative 0;	Mismatches 150;	Indels 0;	Gaps 0;
QY	150	CGGTACTCTCAACAAGAATGTGTGTCGGGAAGGGTGTGAACCTCGAATTTTTCGATCAGCA	209		
DB	164	CSCGATCAAGACGCCGACCTGCTGTCGGCGATGCGCTGAAGCTCAACATCTCTCGCGCAGCA	223		
QY	210	AGATCGCNAATCAGACCCGATGGATCCGGATCTCAACTACGGTGAAGNAGTAGTCAGCAAGCT	269		
QY	224	CAGCCCGGCTCAGACACCGCGCTGGGCGAGGCTATGGACTACGTCGGCGGCGCTCTCAAGTCGCT	283		
QY	270	CGATTCGACCGCGCTGAAGNAGNATCTCCACGCTGTGATGACCGGACAGATCGATCGTGC	329		
DB	284	CGACTACGAGGCGCTCAGAGAGAGCTCAGCGCTCGATATACGCTGATATACATCTCTCTAT	343		

330 GCGCGCTGACTGGGGGCACTACGGCGGTTTGATGATCCGTATGGCTTGGCACTCCGGTGG 389

Db 344 GCWGGCGACTTTGGCCACTACGGCGGTCTGTTTCATCCGCAATGGCCTGGCACAGGCCGG 403

390 CACCTACCGTATGCTGATGCCCGTGGGGCGGTGTACGGGAAGCAGCGCTTGCACC 449

Db 404 CACGTACCGAGTTCACGACGGACGGCGAGGTGGTGGAGAGGCCAGCAACGCTTCGACC 463

QY 450 GCTCAACTCCTGGCCGGGCAACGTCAGCCTGGATAAAGCGCGCGCTCTGCTGTGGCCGAT 509

D**b** 464 GGTCAACAGCTGGCGCGGACAAATGTCAGCCTCGACAAGGCCCGTGGGCTGCTGTGGCCCAT 523

510 CAAGAAGACWACGGCAACAAAATCAGCTGGGCAGACCTGATCATTCCTGGCTGGCACCGT 569

Db 524 CAAGCAAAAGTACGGCAACAAGATCTCGTGGCGACCTCATGATCTCTGGCGGGCAACGT 583

570 GGCATTATGAGTCCATGGGCTTACCTGGTTACGGCTTCTCTTTGGCCGGCTCGATA 625

Db 584 GSCCCTCGAGTCCATGGGTTTCCAGAGGGGGGCTTCTCCGAGGGCGTCCGACA 639

RESULT /  
BQ751387

LOCUS BQ/5138/ 606 bp mRNA linear EST 18-JUL-2002  
DEFINITION EST631950 DSCT Colletotrichum trifolii cDNA clone pDSC4-87, mRNA

sequence.  
ACCESSION BQ751387

VERSION BQ751387.1 GR:21906792  
KEYWORDS EST.

SOURCE	ORGANISM
Colletotrichum trifolii	Colletotrichum trifolii
Colletotrichum trifolii	Colletotrichum trifolii

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;

mitosporic Phyllachoraceae; Colletotrichum.  
REFERENCE 1 (bases 1 to 606)

**AUTHORS**  
Samac, D.A., Dickman, M., Town, C.D., Van Aken, S., Utterback, T.,  
Cheung, F. and Fraser, C.M.

TITLE	ESTs from mycelia of <i>Colletotrichum trifolii</i> race 1
JOURNAL	Unpublished
1	

Other\_ESTs: EST631951  
Contact: Deborah A Samac

Contact: Deborah A. Sunde  
Department of Plant Pathology  
University of Minnesota

University of Minnesota  
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA  
Tel.: 612 535 1242

tel: 012 923 1243  
Fax: 651 649 5058

Email: [debbys@puccini.cit.umn.edu](mailto:debbys@puccini.cit.umn.edu)  
TIGR sequence name: MTSAD87TK More information is available at:

www.medicago.org  
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).

FEATURES	Location/Qualifiers
source	1. 606

/organism="Colletotrichum trifolii"

```

/mol_type="mrna"
/strain="ace 1" 5466"
/clone="pBSG24-87"
/tissue="mycelia"
/dev_stage="Young, actively growing mycelia (3 days after
inoculation) grown in liquid culture (cutin minimal medium
containing 2%glucose)."
/lab_host="DH5alpha"
/clone_lib="DSCRTBLuescript SK+, Site-1, EcoRI, Site-2;
EcoRI; isolate; 2ap2 cDNA was prepared from polyA+
enriched RNA. The cDNA was ligated into Lambda stII from
Stratagene and packaged using GigaPack packaging extracts.
An aliquot of the amplified library was used to transduce
E. coli X1090 and phage DNA was purified from a liquid
lysate. The cDNA inserts were gel purified after EcoRI
digestion and ligated into pBluescript SK+. Aliquots of
the recombinant library were transformed into E. coli DH5alpha which
were plated onto medium with X-gal for selection of
recombinants."
BASE COUNT      124 a 214 c 172 g 96 t
ORIGIN
Query Match      10.4%; Score 234.8; DB 13; Length 606;
Best Local Similarity 55.8%; Positives 146; Negatives 18;
Matches 337; Conservative 0; Mismatches 152; Indels 1; Gaps 1;
QY 150 CGGTACTTCCACAAAGATTGGTGGCGGAAGGTTGAACCTGGATATTTCATCAGCA 209
DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 117 CGGCATCAGAAACCGGACTGTGGCGGATGGCGGTGAAGCTCAACATCTCCGCCAGCA 176
QY 210 AGATCGAATCAGACGCGATGGATCCGATTCACATCGTGAGAGGTAGCGAGCT 269
DB 177 CAGCGCGTCCACCAACCCGCTGGCGCAGACTTGAATAGCTCCGCGCTTCAAGTCCCT 236
QY 270 GATTTCACGCGGTGAAGAAAGATGTCCACGGGTTGATGACCGATAGCCAGAGTGGTG 329
DB 237 GCACTAGAGGCGCTCAGAGAGAGCTCAACGGCCCTGATGACCGATCCCGAGACTGGT 296
QY 330 GCGCGTGAATGCGGCGCTACGGCGGTTTGATGATCGGTATGCTTGGCATCGGTGG 389
DB 297 GCGTCCGACTTTGGCCACTACGGCGCTCTGTCATCCGATGCGCTTGGCAGCGCGCG 356
QY 390 CACCTCATCTATTCTGATGGCGGTGG-GGCGGGTGTACCGAAGCCAGCGCTTTTGCAC 448
DB 347 CACATCGAGGTTCAACGACGAGCGGCTGAGGCGGTGGAGCGGCCAACACGCTTGCAC 416
QY 449 CGCTCATCTCCGCGACGACGCTGAGCTGGATTAAGCGCGCGCTCTGCTGGCGGCA 508
DB 417 CGCTCAACAGCTGGCGCGCAATATGTCAGCTCGACAGCGCGCTCTGGCTGTGGCGCA 476
QY 509 TCAAGAGAGTACGACCAACAAATACGTGGCGAGACCTGATGATTCGTGGTGGCAGCG 568
DB 477 TCACGAAGAGTACGCGCAACAGATGTCGTGGCGGCACTCATGATCTCTGGCGCGCAG 536
QY 569 TGGCTATGATGCTTCCGCTTACGCTGCTACGGCTCTCTTCTTGGCGCGCTGCAATTT 628
DB 537 TGGCGCTTGATGCTTCCAGAGCGCGGCTCTCTCGGAGCGCGCTCGCGAGACT 596
QY 629 GGGAAACCGCA 638
DB 597 GGGAGGCGCA 606

```

## RESULT 8

```

AW179968      729 bp  mRNA      11near  EST 17-NOV-1999
LOCUS      Mga020f Mga Library Mycosphaerella graminicola cDNA clone Mga0020
DEFINITION      Similar to catalase-peroxidase, mRNA sequence.
ACCESSION      AW179968
VERSION      AW179968.1 GI:6447163
KEYWORDS      EST.

```

## SOURCE

ORGANISM Mycosphaerella graminicola  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et  
 Mycosphaerellales; Mycosphaerellaceae; Mycosphaerella;

## REFERENCE

1 (bases 1 to 729)

## AUTHORS

Keon,J.P.R., Bailey,A.M. and Hargreaves,J.A.

## TITLE

A group of expressed cDNA sequences from the wheat fungal leaf

blotch pathogen, Mycosphaerella graminicola (Septoria tritici).

Fungal Genet. Biol. 29 (2), 118-133 (2000)

## JOURNAL

2004020

## COMMENT

Contact: Hargreaves, JA  
 Cell Biology Department  
 IACR-Long Ashton Research Station  
 Long Ashton, Bristol, BS41 9AP, UK  
 Tel: +44(0)1275 392181  
 Fax: +44(0)1275 394281  
 Email: john.hargreaves@bsrc.ac.uk  
 Submitted: 13 Aug 2000 Std Error: 0.00  
 Seq Primer: 13

## FEATURES

source

1..729

/organism="Mycosphaerella graminicola"

/mol\_type="mrna"

/strain="Strit"

/db\_xref="Genbank:54734"

/clone\_lib="Mga Library"

/note="Vector: pSPORI; Library constructed from cultures

utilizing ammonium ions as a source of nitrogen"

BASE COUNT 159 a 217 c 203 g 150 t

ORIGIN

Query Match 10.3%; Score 233.2; DB 9; Length 729;

Best Local Similarity 55.3%; Positives 146; Negatives 18;

Matches 343; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 151 GGTACTTCCACAAAGATTGGTGGCGGAAGGTTGAACCTGGATATTTCATCAGCAA 210

DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

DB 92 GTATCCCGCACACGGGACTGTGGCGAATGAGCTGAACCAAGATCTCGCGCAGCAC 151

QY 211 GATCGCAATACAGACCGATGATCGGATTCACATCGCTGAGCAAGTACGACGCT 270

DB 152 ACGCTGTCTACCGACCATTCGCGCAAGCTGCTACCTACCCAGCATTCAGAGCTTG 211

QY 212 GATTTCGACCGCTGAAGAAAGATCTCCACGCTGTGATGACCATAGCAGCAAGTGTGG 330

DB 271 GATTACATACGCGGCTCAGAGAGCTGACACGCTCATGACCGAGTCTGGGAGACTCT 271

QY 331 CCGCTGTACTGGGCGCTACGCGGCTGTGATGATCGCTGATGCTGTGCGCTCGCTGG 390

DB 272 CCAGCAGACTTTGGCCACTACGCTGTTCTCTGTCGCGCATGBCATGGCACAGCGCT 331

QY 391 ACCTACGCTATTGCTGATCGCGCTGGGGCGGTGTACCGAAGCCAGCGCTTTGCACCG 450

DB 332 ACTACGCTGCTGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 391

QY 451 CTCACACTCTCTGGCGGACAGCTGACGCTGGTAAAGCGCGCTCTGCTGTGGCGGATC 510

DB 392 CTCACACTCTCTGGCGGACAGCGCATCTCGACAGGCTCGCGGCTGTGTGGCGCAATC 451

QY 511 AAGAGAGAGTACGCAACAAATACGCTGGGCGAGCTGATGATTCGCTGGCGCACCGTG 570

DB 452 AAGCAAGGTACGGAACAAAGATCATGATGGCGGCACTTCTCTCTCTCTGAAATGTC 511

QY 571 GCTTATGATGCTCCATGGGCTACCTGCTTACGCGCTCTCTTGTGGCGCGGCTCAATGTC 630

DB 512 GCAATGATGCTTATGGGCTGCGCCACACTTTGTTTTCGCGGCTGTGGCGGACACATGG 571

QY 631 GACCCGGAAGATATCTACTGGGGGTGACGAAAGAGTGGCTGG 676

DB 572 GAGCGGATGATCTCCCTTACTGGGGGTGTGAGCCACTGTGTTGG 617

Db		353	GCGTACTACGTGGAACCCGGAAAGGCC - GAAGCGCAGCGGTGAACCCCTGTCAGCGCGGC 411
QY		812	AGCAGGTACTCTGAACATCTTGCCGGCGGTATGGCGGTGAACCAAGCAAAAAAAGCCAGCGCCCTCA 871
Db		412	AGC-GATTGGTGGCATCTTGCGCAACAATGGCGCATGAACAGTGAAGAAGAACCGTCGCGCTGA 470
QY		872	CAGCTGCGCGCACACAGCTGCGGTGTTTCTGCGGTATGCGGTATGCGATTAAGCATCGCTGTAGCGG 931
Db		471	TTCGACGGCGCACACCTCCGCAAGCATTCACGGCGCGGTGAAGCCACCACCGTGGGTGA 530
QY		932	GTGACCCAAAAGAGCTGTGACGTCTGAAACACAGCGGCTTAGTGTGGGCGAACACCCCAACATGC 991
Db		531	CCGAGCCCGAGAGCGCTCAACGATTAAAMRCAACAGCTGGGCTGGGCGCAGCAGCGCAGCGCA 990
QY		992	AGGCGCAAGCAACAGACGCGGTGACCTGCGGTGATGCGAGGTGCTGTGGCACACACAGCGG- 1050
Db		591	CGCGCGTGGTGGCGTGATGCCATFACCTCGCGGTCTGSAACAGTCATCTGGTCACAAACCCCG 650
QY		1051	ACGAAATTCGATATGGGCTATVTTTCGACATCTGCTGTTCGGTCACTATGGGAGCTGAAAAAG 1110
Db		611	AACCCAGTGGAGCACTACTCTTCTTCGAGAACCTGTTCAATAACGAATGGGTACAGACCCGC 710
QY		1111	AGTCCGCGCGCGCCACAGTATGGGA 1137
Db		711	AGCGCGCAGCGSGCANTCTGTTGAA 737

  

RESULT 10	BMB65076	663 bp mRNA linear EST 06-MAY-2003
LOCUS	mgap008x018f.b Magnaporthe grisea Ap Uni-zap XR library Magnaporthe	
DEFINITION	Magnaporthe grisea clone mgap008x018f 5', mRNA sequence.	
ACCESSION	BMB65076	
VERSION	BMB65076.2	GI:30390127
KEYWORDS	EST.	
SOURCE	Magnaporthe grisea (anamorph: Pyricularia grisea)	
ORGANISM	Magnaporthe grisea	
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;	
AUTHORS	Chenomyces-like fungus; Magnaporthe sensu lato; Magnaporthe grisea; Magnaporthe	
	1 (bases) to 663	
	Ebbel,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P., Lu,G., Bhatterai,	
	K. and Dean,R.A.	
TITLE	Expressed sequence tags from the rice blast fungus, Magnaporthe	
JOURNAL	grisea	
COMMENT	Unpublished	
	On Mar. 7, 2002 this sequence version replaced gi:1923758.	
	Department of Plant Pathology & Microbiology	
	Texas A&M University	
	Peterson Bldg, MS2132, College Station, TX 77843-2132, USA	
	Tel.: 979 845 4831	
	Fax: 979 845 6483	
	Email: d-ebbel@tamu.edu	
	Chromatogram file of this sequence is available, see contact person	
	(Neurospora crassa) g13128921 catalase-peroxidase	
PCR PRIMERS	>g128921..288 48-77	
FORWARD:	T3 primer	
BACKWARD:	T7 primer	
Plate:	mgap008 row: D column: 18	
Seq primer:	T3.	
Location/Qualifiers		
/organism=	"Magnaporthe grisea"	
/mol_type=	"mRNA"	
/strain=	"70-15"	
/db_xref=	"taxon:148305"	
/clone=	"mgap008x018"	
/sex=	"Mat1-2 hermaphrodite"	
/cell_type=	"Appressorium"	
/notes=	"Magnaporthe grisea Ap Uni-zap XR Library"	
/note=	"Vector construct: pSP+::Sic1::ERT::SUC2, XhoI"	
/note=	"Unidirectional cloning. Eppri has T3 primer and"	

  

FEATURES	source
----------	--------





[illegible]

**BASE COUNT**



Qy	2026	GATTTGGGAGACAGCTGGAGACCGGTAGTGTACAGGCC---	TACGAATTCGGCAGCGC	2082
Db	346	GACATCAGCAGCGAGTGGGAAGCCCTCGGCAACGGCACTGTT	CGAGGATATTGACGCG	287
Qy	2083	AAGACCGGGTGGCGTGAAGTGCAGACCGCTCGGGGGTGGAT	TGTGGTTTCCCAACTGC	2142
Db	286	AAGACGGGCGACAAAGAGTGGCATGGTACCCCTGTGCATCT	TAGCTCTTGGATCACAAGCT	227
Qy	2143	CTACTGGCGTCTTACGCGAGATGTACGCCGAGAGATTCGT	TCGACG	2202
Db	226	GACCTTCGGCGCGACCTGACGAAGATATGCACTCGGGCT	ACGCGACGCTGGTTCGC	167
Qy	2203	GACCTTCGGCGCGCTGGACCAAGATGATGACACGCGACCG	TGTGCAGCT	2252
Db	168	GACCTTATCGCTGCTGGACCAAGGTGATGACCTTGACCGAT	TTTGTATGT	17

Search completed: October 8, 2003, 15:29:18  
Job time : 4695.37 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (C) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: October 7, 2003, 19:31:22 ; Search time 60.8231 Seconds  
(Without alignments)  
1985.081 Million cell updates/sec

title: US-09-884-889-6

Perfect score: 4058

Sequence: 1 MNASADDLHSLQRCRAF.....RDFVAANTKVHNRDFVAS 753

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_13Jun03.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score distribution, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4058	100.0	753	19 AAW33809	Alcaligenes (Deley)
2	4058	100.0	753	23 ABG33040	Alcaligenes catala
3	2820	69.5	723	23 ABG78212	Amino acid sequenc
4	2290.5	56.4	745	19 AAW33810	Microscilla furves
5	2044.5	50.4	740	18 AAW26596	Mycobacterium tube
6	2044.5	50.4	740	20 AAW95598	M. tuberculosis ca
7	2044.5	50.3	740	20 AAW95598	Mycobacterium tube
8	2044.5	50.3	740	20 AAW95598	M. tuberculosis ca
9	1890	46.6	735	14 AAR43670	M. tuberculosis H37

10	1890	46.6	735	20 AAW78363	M. tuberculosis kat
11	1547	38.1	731	11 AAR08205	Heat resistant per
12	1052	19.3	273	22 ABG08923	Novel human diago
13	780.5	19.3	273	22 ABG08923	Novel human diago
14	519	12.8	1550	22 ABG09297	Novel human diago
15	398	9.8	201	22 ABG08828	Novel human diago
16	272.5	6.7	231	23 ABG05661	Hordeum vulgare pe
17	270	6.7	246	23 ABH92257	Herbicidally activ
18	269	6.6	334	21 AAG44544	Arabidopsis thalia
19	269	6.6	426	21 AAG44544	Arabidopsis thalia
20	269	6.6	426	21 AAG44544	Arabidopsis thalia
21	269	6.6	452	21 AAG44542	Arabidopsis thalia
22	267	6.6	250	21 AAG09702	Arabidopsis thalia
23	267	6.6	250	21 AAG16805	Arabidopsis thalia
24	267	6.6	250	21 AAG26453	Arabidopsis thalia
25	267	6.6	250	21 AAG43523	Arabidopsis thalia
26	267	6.6	250	21 AAG45582	Arabidopsis thalia
27	267	6.6	250	21 AAG45589	Arabidopsis thalia
28	267	6.6	250	21 AAG45589	Arabidopsis thalia
29	267	6.6	250	23 ABG90303	Herbicidally activ
30	267	6.6	254	21 AAG43522	Arabidopsis thalia
31	267	6.6	297	21 AAG26452	Arabidopsis thalia
32	267	6.6	297	21 AAG45588	Arabidopsis thalia
33	265	6.5	215	21 AAG09703	Arabidopsis thalia
34	265	6.5	215	21 AAG16806	Arabidopsis thalia
35	265	6.5	215	21 AAG43524	Arabidopsis thalia
36	265	6.5	215	21 AAG43524	Arabidopsis thalia
37	265	6.5	215	21 AAG45583	Arabidopsis thalia
38	265	6.5	215	21 AAG45586	Arabidopsis thalia
39	265	6.5	215	21 AAG45590	Arabidopsis thalia
40	260.5	6.4	347	21 AAG24504	Arabidopsis thalia
41	260.5	6.4	348	21 AAG05527	Arabidopsis thalia
42	260.5	6.4	353	21 AAG05526	Arabidopsis thalia
43	260.5	6.4	353	21 AAG05526	Arabidopsis thalia
44	260	6.4	279	23 ABG32661	Herbicidally activ
45	259.5	6.4	348	21 AAG52697	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1  
AAW33809  
ID AAW33809 standard; Protein; 753 AA.  
AC AAW33809;  
XX 18-NOV-1998 (first entry)  
XX Alcaligenes (Deleya) aquamarinus catalase-6ACA2.  
DE Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;  
KW pasteurisation.  
XX Alcaligenes aquamarinus.  
XX W09800526-A1.  
XX 08-JAN-1998.  
XX 03-JUL-1997; 97WO-US16513.  
XX 03-JUL-1996; 96US-0674887.  
PR (RECO-) RECOMINANT BIOANALYSIS INC.  
PI Adhikary RS, Robertson DE, Sanyal I;  
DR WPI; 1998-086953/08.  
DR N-PSDB; AAY06554.  
XX New bacterial catalases, related nucleic acid vectors and  
XX transformed cells - used as oxidising agents and for detecting or

destroying hydrogen peroxide, e.g. in biosensors  
 Claim 1: Fig 1; 35pp; English.

The present sequence is of the Alcaligenes equimarinus catalase-64CA2.  
 Catalase-64CA2 may be used to catalyze oxidation reactions such as epoxidation or hydroxylation. The enzyme can also be used to detect or destroy hydrogen peroxide, e.g. in connection with glycolytic acid production, biosensors, contact lens cleaning, pulp/paper bleaching and production of dairy products. Antibodies raised against catalase-64CA2 can be used to detect the enzyme. Fragments of the catalase-64CA2 can be used to detect the enzyme. Fragments of the gene encoding Catalase-64CA2 can be used to identify related sequences.

Query Match 100.0%; Score 4056; DB 19; Length 753;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNNASADLLSSLOQRCRAFPVLPSPHRAIRERAMSKCPVHGCGTGTGTGSKNDWPE 60  
 DB 1 MNNASADLLSSLOQRCRAFPVLPSPHRAIRERAMSKCPVHGCGTGTGTGSKNDWPE 60  
 QY 61 GLINDILHQDQKSDPMDPNTREVKRLDFDALKDVKHALMTDSQWMPADMGHTGGL 120  
 DB 61 GLINDILHQDQKSDPMDPNTREVKRLDFDALKDVKHALMTDSQWMPADMGHTGGL 120  
 QY 121 MTRMAHISAGTYRIADRGCGGTGSGRFAPLNSPDNYSIDKARLLAPTKKYGKISW 180  
 DB 121 MTRMAHISAGTYRIADRGCGGTGSGRFAPLNSPDNYSIDKARLLAPTKKYGKISW 180  
 QY 181 ADMATLACTVAYESKGLPAFGSGFGRVDIWEPEKDIYWGDEKEMAPSDERTGVNKPET 240  
 DB 181 ADMATLACTVAYESKGLPAFGSGFGRVDIWEPEKDIYWGDEKEMAPSDERTGVNKPET 240  
 QY 241 MENFLAAMVOMGLIYVNPVNGHPDPURTAQOVLTFEARMANDEKTAALTAGGHTVGC 300  
 DB 241 MENFLAAMVOMGLIYVNPVNGHPDPURTAQOVLTFEARMANDEKTAALTAGGHTVGC 300  
 QY 301 HGNHNASALAPDKASDVENOGGLGNGPNQKASNAVTSLEGAMNTPTKMGYFDL 360  
 DB 301 HGNHNASALAPDKASDVENOGGLGNGPNQKASNAVTSLEGAMNTPTKMGYFDL 360  
 QY 361 LFQYVNEKKSFGAHHEPTDCKENKVDASDSIRHNPIMTDADMAIKVNPYTRAC 420  
 DB 361 LFQYVNEKKSFGAHHEPTDCKENKVDASDSIRHNPIMTDADMAIKVNPYTRAC 420  
 QY 421 EKPMADPEYKFKTAKAMEKLTHERLDGPKSVRYGPEVPADLIINQDIPAGNTDYCEVY 480  
 DB 421 EKPMADPEYKFKTAKAMEKLTHERLDGPKSVRYGPEVPADLIINQDIPAGNTDYCEVY 480  
 QY 481 KQRTAGSLGSIEMVSTAMDSATYRSGDMRGANGARIELAPQWENQNEPERLAKVLS 540  
 DB 481 KQRTAGSLGSIEMVSTAMDSATYRSGDMRGANGARIELAPQWENQNEPERLAKVLS 540  
 QY 541 VTQISDGTASADIVLAGSVGTAKAAKYDYVRVPLKRGCGATAEKWDADSNPL 600  
 DB 541 VTQISDGTASADIVLAGSVGTAKAAKYDYVRVPLKRGCGATAEKWDADSNPL 600  
 QY 601 BFLADGFRNKKYVYVPEMLDRAQWGLTGPMTVLLGNGVLTGNTGKGVFT 660  
 DB 601 BFLADGFRNKKYVYVPEMLDRAQWGLTGPMTVLLGNGVLTGNTGKGVFT 660  
 QY 661 DCEGOLTNDFVNLTKMNGSKPVGSNAVETROKTKGAVKMTASRDLVFGSNLSJASY 720  
 DB 661 DCEGOLTNDFVNLTKMNGSKPVGSNAVETROKTKGAVKMTASRDLVFGSNLSJASY 720  
 QY 721 EYVAQDNGEKTYRDFVAAMTKVNAHDRDVAS 753  
 DB 721 EYVAQDNGEKTYRDFVAAMTKVNAHDRDVAS 753

RESULT 2  
 ABG33040 standard; Protein; 753 AA.  
 AC ABG33040;  
 DT 22-NOV-2002 (first entry)  
 DE Alcaligenes catalase protein 64CA2.  
 KW Gene; ss; marine bacterla; catalase; hydrogen peroxide; enzyme.  
 OS Alcaligenes aquamarinus.  
 XX US2002102680-A1.  
 XX 01-AUG-2002.  
 XX 19-JUN-2001; 2001US-0864889.  
 PR 03-JUL-1996; 96US-0674887.  
 PR 16-OCT-1997; 97US-0951844.  
 PR 05-OCT-1999; 99US-0412347.  
 XX (ROBE/) ROBERTSON D E.  
 PA (SANT/) SANTAL I.  
 XX (ADHL/) ADHIKARI R.  
 PI Robertson DE, Sanyal I, Adhikari R;  
 DR WPI; 2002-690613/74.  
 DR N-PDB: ABS53944.  
 XX New purified catalase polypeptide useful for catalyzing the breakdown of hydrogen peroxide, for modifying small molecules, and for generating antibodies which bind to the polypeptide.  
 PS Claim 22: Fig 5; 44pp; English.  
 XX This invention relates to the DNA and protein sequences of a novel purified catalase polypeptide. The protein of the invention or its homologues is useful for catalyzing the breakdown of hydrogen peroxide. The polypeptide of the invention is useful for modifying small molecules, by mixing the protein or its fragments with the small molecule to produce a modified small molecule. The protein of the invention is also useful for catalyzing biochemical reactions, for hydrolyzing glycosidic linkages and for generating antibodies which bind specifically to the protein. The nucleic acid sequences of the invention is useful as a probe to determine whether a biological sample, an organism from which the DNA was isolated, contains the gene for an organelle from which the DNA was isolated. The polypeptide of the invention is useful as an antibody that binds the cellulase isolating related nucleic acids. An antibody that binds the cellulase protein of the invention is useful in immunodiffusion chromatography procedures to isolate or purify the protein or for detection of protein expression in a biological sample. The cellulase protein of the invention is heat stable, is heat resistant, and is able to denature and regain activity after exposure to temperatures of from 60 to 100 degrees Celsius. The present sequence represents the Alcaligenes equimarinus catalase protein of the invention.

Query Match 100.0%; Score 4056; DB 23; Length 753;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNNASADLLSSLOQRCRAFPVLPSPHRAIRERAMSKCPVHGCGTGTGTGSKNDWPE 60  
 DB 1 MNNASADLLSSLOQRCRAFPVLPSPHRAIRERAMSKCPVHGCGTGTGTGSKNDWPE 60  
 QY 61 GLINDILHQDQKSDPMDPNTREVKRLDFDALKDVKHALMTDSQWMPADMGHTGGL 120  
 DB 61 GLINDILHQDQKSDPMDPNTREVKRLDFDALKDVKHALMTDSQWMPADMGHTGGL 120

121 MIRNAHESACTYRIADRGCGGTGSGOREPAPLNSPONSIDKARLLHDIKKYKNSLW 180  
 Db 121 MIRNAHESAGTYRIADRGCGGTGSGOREPAPLNSPONSIDKARLLHDIKKYKNSLW 180  
 Qy 181 ADMLIAGTVAYESMGIPATGFSFGVNDIWEPEKDIYMGDEKWLAPSDERYGDVNRKPT 240  
 Db 181 ADMLIAGTVAYESMGIPATGFSFGVNDIWEPEKDIYMGDEKWLAPSDERYGDVNRKPT 240  
 Qy 241 MENPLAIVOMGLIYVNEPEGVNGHDPFETAQOULETFARNAWDEKTAALTAGHTVNC 300  
 Db 241 MENPLAIVOMGLIYVNEPEGVNGHDPFETAQOULETFARNAWDEKTAALTAGHTVNC 300  
 Qy 301 HONGMSALAPDKASDVENQGLWGNPNMQKASNAVTSIGEGAWTPTFKDMGYEDL 360  
 Db 301 HONGMSALAPDKASDVENQGLWGNPNMQKASNAVTSIGEGAWTPTFKDMGYEDL 360  
 Qy 361 LGCTYMWELKSPAGAHHEPDIKKENKPYDASDPSIRNEIMTDAOMAIVNPTVTAIC 420  
 Db 361 LGCTYMWELKSPAGAHHEPDIKKENKPYDASDPSIRNEIMTDAOMAIVNPTVTAIC 420  
 Qy 421 EKEMADPEYFKTEAKAFKLRHDIHGPKSYIGVEPAEDLWIDPEIPAGNTDICEPV 480  
 Db 421 EKEMADPEYFKTEAKAFKLRHDIHGPKSYIGVEPAEDLWIDPEIPAGNTDICEPV 480  
 Qy 481 KOKTAGSGLSISEMVSANDSARTYRGSDMRGANGARILAPENOMQNEPERLAKVLS 540  
 Db 481 KOKTAGSGLSISEMVSANDSARTYRGSDMRGANGARILAPENOMQNEPERLAKVLS 540  
 Qy 541 VYEQI SAUTGASTADIVILAGSYGTEKAAAGTDRVPFIKRGDATABEMTDADSFAPL 600  
 Db 541 VYEQI SAUTGASTADIVILAGSYGTEKAAAGTDRVPFIKRGDATABEMTDADSFAPL 600  
 Qy 601 EPLADGRFRNMOKKEYVYKPEMLIDRAQLMGLTGPEMTVLLGGMVLTGNTGSKRGVPT 660  
 Db 601 EPLADGRFRNMOKKEYVYKPEMLIDRAQLMGLTGPEMTVLLGGMVLTGNTGSKRGVPT 660  
 Qy 661 DCEQULINDPFWLIDMGNSMKPVGSNAYEIRDRKTAIVKVTASRVOLVFGSNLSLRSTA 720  
 Db 661 DCEQULINDPFWLIDMGNSMKPVGSNAYEIRDRKTAIVKVTASRVOLVFGSNLSLRSTA 720  
 Qy 721 EYVAQDNGCEKFYRDFVAAATKYNNADRFDVAS 753  
 Db 721 EYVAQDNGCEKFYRDFVAAATKYNNADRFDVAS 753  
 RESULT 3  
 ABB78212  
 ID ABB78212 standard; Protein; 723 AA.  
 XX AC ABB78212;  
 XX DT 25-NOV-2002 (first entry)  
 XX XX Amino acid sequence of Agrobacterium tumefaciens catalase kAta gene.  
 XX XX Catalase; kAta; reactive oxygen species; ROS; ROS-inducible promoter;  
 XX XX H202 level; antioxidant.  
 XX OS Agrobacterium tumefaciens.  
 XX XX WO200263032-A1.  
 XX PD 15-AUG-2002.  
 XX XX 05-FEB-2002; 2002WO-SG00018.  
 XX XX 05-FEB-2001; 2001US-266657P.  
 XX XX (PANS/) PAN S Q.  
 XX Pan 50;  
 XX

DR NPI: 2002-643421/69.  
 XX N-PSDB; ABB78539.  
 PT Determining the ability of a compound to remove a reactive oxygen  
 PT species (ROS), useful in pharmaceutical and nutraceutical fields and  
 PT measuring activities of antioxidants, comprises providing a cell having  
 PT a ROS-inducible promoter -  
 XX  
 XX Disclosure; Page 63-67; 70pp; English.  
 CC The present sequence represents Agrobacterium tumefaciens catalase  
 CC (kAta). kAta has a promoter which is induced by a reactive oxygen species  
 CC (ROS). The kAta gene can be used in the method of the invention. The  
 CC specification describes a method for determining the ability of a  
 CC compound to remove a ROS. The method comprises providing a cell having  
 CC a ROS-inducible promoter operably linked to a reporter gene that is  
 CC heterologous to the promoter, and exposing the cell to a compound  
 CC capable of inducing the promoter. The method also includes determining  
 CC the ability of a compound or gene product to remove ROS and determining  
 CC the level of the ROS H202 inside a cell, to properly assess and label  
 CC antioxidant products for use as food supplements, natural products or  
 CC drugs. The method may also be used in selecting for new or improved ROS  
 CC scavengers.  
 XX  
 XX Sequence 723 AA;  
 Qy Query Match 69.5%; Score 2820; DB 23; Length 723;  
 Best Local Similarity 71.0%; Pred. No. 6.8e-236;  
 Matches 508; Conservative 86; Mismatches 121; Indels 0; Gaps 0;  
 Qy 37 SKCFVYHGGTSTGTGSKNDWMPGELNDILHQDQRKSDPMDPNTREYRKILDFALK 96  
 Db 37 SKCFVYHGGTSTGTGSKNDWMPGELNDILHQDQRKSDPMDPNTREYRKILDFALK 96  
 Qy 97 KDVIALLTDSEWPADEWBGYGLMTIRAHISAGTYRIADRGSGGTGSGOREPAPLNSWD 156  
 Db 97 KDVIALLTDSEWPADEWBGYGLMTIRAHISAGTYRIADRGSGGTGSGOREPAPLNSWD 156  
 Qy 68 ADLALMTDSEWPADEWBGYGLMTIRAHISAGTYRIADRGSGGTGSGOREPAPLNSWD 127  
 Db 68 ADLALMTDSEWPADEWBGYGLMTIRAHISAGTYRIADRGSGGTGSGOREPAPLNSWD 127  
 Qy 157 NVSLDKARLLHDIHGPKSYIGVEPAEDLWIDPEIPAGNTDICEPV 216  
 Db 157 NVSLDKARLLHDIHGPKSYIGVEPAEDLWIDPEIPAGNTDICEPV 216  
 Qy 128 NVYDQKGRLLHDIHGPKSYIGVEPAEDLWIDPEIPAGNTDICEPV 187  
 Db 128 NVYDQKGRLLHDIHGPKSYIGVEPAEDLWIDPEIPAGNTDICEPV 187  
 Qy 217 YNCDKEKWLAPSDERYGDVNRKPTVLAAYOMGLIYVNEPEGVNGHDPFETAQOULEPT 276  
 Db 217 YNCDKEKWLAPSDERYGDVNRKPTVLAAYOMGLIYVNEPEGVNGHDPFETAQOULEPT 276  
 Qy 188 YNCDKEKWLAPSDERYGDVNRKPTVLAAYOMGLIYVNEPEGVNGHDPFETAQOULEPT 247  
 Db 188 YNCDKEKWLAPSDERYGDVNRKPTVLAAYOMGLIYVNEPEGVNGHDPFETAQOULEPT 247  
 Qy 277 FARNAWDEKTAALTAGHTVNCNGCHGNASALAPDKASDVENQGLWGNPNMQKASN 336  
 Db 277 FARNAWDEKTAALTAGHTVNCNGCHGNASALAPDKASDVENQGLWGNPNMQKASN 336  
 Qy 248 FARNAWDEKTAALTAGHTVNCNGCHGNASALAPDKASDVENQGLWGNPNMQKASN 307  
 Db 248 FARNAWDEKTAALTAGHTVNCNGCHGNASALAPDKASDVENQGLWGNPNMQKASN 307  
 Qy 337 AVTSGIEGAWTPTFKDMGYEDLHGYNWLKSPAGAHHEPDIKKENKPYDASDPS 396  
 Db 337 AVTSGIEGAWTPTFKDMGYEDLHGYNWLKSPAGAHHEPDIKKENKPYDASDPS 396  
 Qy 308 TVYSIEGAWTPTFKDMGYEDLHGYNWLKSPAGAHHEPDIKKENKPYDASDPS 367  
 Db 308 TVYSIEGAWTPTFKDMGYEDLHGYNWLKSPAGAHHEPDIKKENKPYDASDPS 367  
 Qy 397 IIRNIMTDAOMAIVNPTVTAICEREMADPEYFKTEAKAFKLRHDIHGPKSYIGVE 456  
 Db 397 IIRNIMTDAOMAIVNPTVTAICEREMADPEYFKTEAKAFKLRHDIHGPKSYIGVE 456  
 Qy 368 IIRNIMTDAOMAIVNPTVTAICEREMADPEYFKTEAKAFKLRHDIHGPKSYIGVE 427  
 Db 368 IIRNIMTDAOMAIVNPTVTAICEREMADPEYFKTEAKAFKLRHDIHGPKSYIGVE 427  
 Qy 457 VPAEDLWIDPEIPAGNTDICEPVYKPEMLIDRAQLMGLTGPEMTVLLGGMVLTG 516  
 Db 457 VPAEDLWIDPEIPAGNTDICEPVYKPEMLIDRAQLMGLTGPEMTVLLGGMVLTG 516  
 Qy 428 VPAEDLWIDPEIPAGNTDICEPVYKPEMLIDRAQLMGLTGPEMTVLLGGMVLTG 487  
 Db 428 VPAEDLWIDPEIPAGNTDICEPVYKPEMLIDRAQLMGLTGPEMTVLLGGMVLTG 487  
 Qy 517 ARILAPENOMQNEPERLAKVLSYIGVEPAEDLWIDPEIPAGNTDICEPV 576  
 Db 517 ARILAPENOMQNEPERLAKVLSYIGVEPAEDLWIDPEIPAGNTDICEPV 576  
 Qy 577 RVPELKGKGRDATABEMTDADSFAPLADGRFRNMOKKEYVYKPEMLIDRAQLMGLTG 636  
 Db 577 RVPELKGKGRDATABEMTDADSFAPLADGRFRNMOKKEYVYKPEMLIDRAQLMGLTG 636  
 Qy 548 AVPEFAGKGRDATABEMTDADSFAPLADGRFRNMOKKEYVYKPEMLIDRAQLMGLTG 607  
 Db 548 AVPEFAGKGRDATABEMTDADSFAPLADGRFRNMOKKEYVYKPEMLIDRAQLMGLTG 607  
 Qy 637 MTVLLGGMVLTGNTGSKRGVPTVLAAYOMGLIYVNEPEGVNGHDPFETAQOULEPT 696  
 Db 637 MTVLLGGMVLTGNTGSKRGVPTVLAAYOMGLIYVNEPEGVNGHDPFETAQOULEPT 696  
 Qy 608 IIVYDQKGRLLHDIHGPKSYIGVEPAEDLWIDPEIPAGNTDICEPV 667  
 Db 608 IIVYDQKGRLLHDIHGPKSYIGVEPAEDLWIDPEIPAGNTDICEPV 667

QY 697 GAVVWTAARVDLVFGSVLSLSTAVYAAVDGNGEKKFVRDVFVAATKTKVHADRDV 751  
 Db 668 GARYSARVDLVFGSVLSLSTAVYAAVDGNGEKKFVRDVFVAATKTKVHADRDV 722

RESULT 4  
 AAW33810  
 ID AAW33810 standard; Protein: 745 AA.  
 XX AAW33810;  
 AC AAW33810;  
 DT 18-NOV-1998 (first entry)  
 DE Microscilla furvescens catalase-53Cal.  
 XX Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;  
 XX pasteurisation.  
 OS Microscilla furvescens.  
 XX  
 XX MO9800526-A1.  
 FN 08-JAN-1998.  
 PD 03-JUL-1997; 97MO-US16513.  
 PR 03-JUL-1996; 96US-0674887.  
 XX  
 XX (RECO-) RECOMBINANT BIOCATALYSIS INC.  
 PA Adhikary RS, Robertson DE, Sanyal I;  
 XX N-PSDB: AAV05555.  
 DR  
 XX New bacterial catalases, related nucleic acid vectors and  
 XX transformed cells - used as oxidising agents and for detecting or  
 XX destroying hydrogen peroxide, e.g. in biosensors

Claim 1; Fig 2; 35pp; English.

The present sequence is of the Microscilla furvescens catalase-53Cal.  
 CC Catalase-53Cal may be used to catalyse oxidation reactions such as  
 CC epoxidation or hydroxylation. The enzyme can also be used to detect or  
 CC destroy hydrogen peroxide. e.g. in connection with glyoxylic acid  
 CC production. Biosensors, contact lens cleaning, pulp/paper bleaching and  
 CC pasteurisation. The sequence is deposited in the EMBL database.  
 CC Catalase-53Cal can be used to screen libraries for detection and  
 CC purification of cells containing the enzyme. Fragments of the gene  
 CC encoding Catalase-53Cal can be used to identify related sequences.

Sequence 745 AA;

Query Watch 56 AA; Score 2200.5; DB 19; Length 745;  
 Best Local Similarity 59.1%; Prod No. 78e-190;  
 Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;

QY 38 GKCPVNHGG---NSTGTSSNKNWPEGLMDLHQDQKSDPMDPDPVREYVKRLDFA 94  
 Db 18 GKCPFTGGSUKOSAGGCTKNRWNPNMLGRLQHSLSDPNDPDPDYAEFKLDLAA 77  
 QY 95 LKIDVHLMTDSQEWKPDGCHYGGLMTRAMHISAGTVRIAGRGGGTGSQRFPALNS 154  
 Db 78 VKIDALMTDSQWPKDYGCHGPTTRAMHISAGTVRIAGRGGGSGSQRFPLNSN 137  
 QY 155 PDNVSLDKARLLKPKKTKYGNKISWADMLTLAGTVATVSGMLPAYGSGRWDIWPEK 214  
 Db 138 PDNANLDKARLLKPKKTKYGRKISWADMLTGNVALETNGFTGFGAGRADVWPEE 197  
 QY 215 DTWGDKEKWLAPSDERY-GDYNKFTWNPENPLAONGLIYNPEGVGHPDPLKTAQV 273  
 Db 198 DYTWGAETWLV--GDKEYEDG---RELNPGLAVONGLIYNPEGVGHPDPLNANDI 251

QY 274 LETFAMANDKTAALTAGGHTVGNCHGNKAS--LAPDPKASDVNCGLWGNMKG 332  
 Db 252 RETFGMNDSEVALIAGGHTFKRGADAEKTVGRPPAAGIEENSLGKNTGTG 311  
 QY 333 KASNAVTSGRGATNTPTKFDGMYFOLLPGVYNELKSPAGAHNRPETDKEKNKDA 392  
 Db 312 HGAUTITSGLEGAMTKPTPMSNNFENLFGYEWELTKSPAGAYOMKPKDAGACTPDA 371  
 QY 393 SDPSIRHNPIMTDADMAIKVNPYRTAICEKFMADPEYFKTKAFAMFKLTHROLDGSKRY 452  
 Db 372 HOPSKSHAPPTIDTLALAMOPDIKISHRYTENPDEPADAFAMAKTKLTHROMGPKVRY 431  
 QY 453 IGPEVPAEDLIMQDPTAGCTDYCS---EYVKOKIAOSGLSISAWVSTANDSARTYGS 508  
 Db 432 LGPEVPEQEDLIMQDPIFVSHPLVDENDIEGLKAKILESGLTVELSTVSTANASASTPNS 491  
 QY 509 DMRGANGARIRLAPQENQGNQENPERLAKVLSVYEQISAD-----TGASTADTVLA 560  
 Db 492 DKRGANGARIRLAPQENQGNQENPERLAKVLSVYEQISAD-----TGASTADTVLA 551  
 QY 561 GSVGIEKAAKAGDVVRPFLKRGDADTAKMTDADSFAPLEPLADGFRNPKKRYTKE 620  
 Db 552 CGAGVEKAAKAGHVEQVPFNPGRADATAGTQDVEAFAPLEPAADGFNYIKPEKYSAE 611  
 QY 621 ENLLDRAQJMLGTGPMNTVLGGMRVLTGNYGTSKGVFTDCBQTLNDFFVNLZDMGNS 680  
 Db 612 ENLVDRQJLSLSENPALVGGNRVLTGTSQGRVFTKPGQLNDFFVNLZDMGNS 671  
 QY 681 WK--PVGSNAYEIRDKTGAVYKASRVLVFGSNLSLLRSYAEVYADDNGCKFVYDFVA 738  
 Db 672 HRAESDKVFEGRDFKTCGEVKNSTGYDLIFGNSNELRALAEVYGCADSEKFPKVPK 731  
 QY 739 ANTKYKNADRPDV 751  
 Db 732 ANAKVVDLDRD 744

RESULT 5  
 AAW26596  
 ID AAW26596 standard; Protein: 740 AA.  
 XX AAW26596;  
 AC AAW26596;  
 DT 25-JAN-2003 (updated)  
 DT 21-JAN-1998 (first entry)  
 XX  
 XX Mycobacterium tuberculosis wild-type catalase-peroxidase.  
 KW Isoniazid resistance; restriction fragment length polymorphism;  
 KW RFP; katG gene; catalase; peroxidase; multiple drug resistance;  
 KW isonicotinic acid hydrazide; INH; tuberculosis; antibiotic;  
 KW tuberculostatic.  
 XX  
 XX Mycobacterium tuberculosis.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FT Key-difference 264  
 FT /note= "mutation to Thr confers INH resistance"  
 FT Miss-difference 315  
 FT /note= "mutation to Thr confers INH resistance"  
 FT Miss-difference 337  
 FT /note= "mutation to Cys confers INH resistance"  
 FT Miss-difference 463  
 FT /note= "mutation to Leu confers INH resistance"  
 XX US5658733-A.  
 PN 19-AUG-1997.  
 PD 07-APR-1995; 95US-0418782.  
 XX 07-APR-1995; 95US-0418782.  
 XX 07-APR-1995; 95US-0418782.





QY 24 VSPHRAIRER---ANSKCPVM-HGNTSTGTSGKOMPGLNLDILHODRKSDDMP 79  
 DB 1 VPSGHPPTHTTTGAASNGCPVGHKKYVGGGQNDMPNRLNKLVLQNPVADPAC 60  
 QY 80 DFNYREVRKLDKDKVHALMTOQWMPADMGHYGGLMIRMAHSAAGTYRIADGRG 139  
 DB 61 AFQAAEVATIDVATLDEEVTWTSOPMPADYGHGGLFIRMAHAAAGTYRIADGRG 120  
 QY 140 GGTGSGRPAPNSPDWNSDKARLLWPKKKYGNKISWADLMLTLAGTYVYSGLPLA 199  
 DB 121 GAGGGRQFAPNSPDWNSDKARLLWPKKKYGNKISWADLMLTLAGTYVYSGLPLA 180  
 QY 200 YGSEGRVQDMPD-EVYWGKAATWL---GDSRY---SGKROLNPLAANVOMGLIYNPEG 259  
 DB 181 FGSGRGRVQDMPD-EVYWGKAATWL---GDSRY---SGKROLNPLAANVOMGLIYNPEG 234  
 QY 260 VNHDPDLPTAQOVLTFPAMANDKTAALPAGHYVGVGNGNKSALADPKASDYE 319  
 DB 235 PGNPDPMANAVDIRFRMANMDVETALVVGHTFGKTHGAGPADLVGPEAPKFL 294  
 QY 320 NQGLGAGNPMQGNKASNVGTIGCAWTTNPTFKDMYFDLLFGYMWELKSPAGAHWE 379  
 DB 295 QMGGLKASSTGTGKDAITSIEVYVWTFKWNDSFLEILGYEMELTSPAGAQYT 354  
 QY 380 PID-TKEKNKVPDASDSIRHNPIMFDAMALNVPYTPACIKCFMADPFFKTFKAK 438  
 DB 355 AKDGAGAGTIPDPFGPGQ--RSPFMALDLSLRVDPIYERITRMLPELADEPAK 412  
 QY 439 FKULHODGPKRSYIGPEVPAEDLIMODIPAGNTDYCEE---VKOKIAQSLISEM 494  
 DB 413 FKULHODGPKRSYIGPEVPAEDLIMODIPAGNTDYCEE---VKOKIAQSLISEM 476  
 QY 495 VSTANDSARTYRGSDMGRCANGARILAPONWOGNEPE-BLAKVLSYVEOI-----SAD 548  
 DB 473 VSTANAAASFRSGDKRGANGRILOPQVGEVNDPDGLKVRTLEETQESFNSAA 532  
 QY 549 TG---ASIADEVILAGSVGTEKAAAGDYVVPFLKRGDATAEMTDSPAPLEPAD 605  
 DB 533 POKITVPSADVLVWGCNATEKAAAGNINITYPTPTDASQEDVSPVLEPAG 592  
 QY 606 GFRNKKKXVYVPEPMLDQALMGLGPMTVLLGSRVYGNVGTGHEVYDCEGO 665  
 DB 593 GFNTYLGKGNPLPNEYLLDQANLLTSLAPMTVLVGLSRLVYLGANKPLPAGVTEASES 652  
 QY 666 LTHDSFVNLTDGNSMK--PVGSNAYEIRDRKTAGVKTASRDVLVFGSNLSLSYAEVY 723  
 DB 653 LTHDSFVNLTDGNSMK--PVGSNAYEIRDRKTAGVKTASRDVLVFGSNLSLSYAEVY 711  
 QY 724 AQDDNGKXKVPDVFVAANTKYNMADREVD 751  
 DB 712 GADDAQKPEVDVFAANDKYNMLDREVD 739  
 RESULT 7  
 ID AAM31343  
 ID AAM31343 standard; Protein, 740 aa.  
 AC AAM31343;  
 XX AAM31343;  
 DT 20-MAR-1998 (first entry)  
 XX Mycobacterium tuberculosis katG protein consensus sequence.  
 KW Tuberculosis; katG gene; isoniazid resistance; INH;  
 KW Isonicotinic acid hydrazide; consensus.  
 OS Mycobacterium tuberculosis.  
 XX US5688639-A.  
 KW 18-NOV-1997.

PF 18-APR-1994; 94US-0228662.  
 PR 18-APR-1994; 94US-0228662.  
 PA (MAYO-) MAYO FOUNDATION.  
 XX Cockerill FR, Kline BC, Uhl JR;  
 XX WPI; 1998-007975/01.  
 PT Determination of Isoniazid sensitivity of Mycobacterium tuberculosis  
 XX strains - by restriction length polymorphism analysis of katG gene  
 XX Example 2; Column 15-20; 18pp; English.  
 PS This protein sequence represents the consensus amino acid sequence of  
 CC the katG gene from Mycobacterium tuberculosis strains H37Rv NC, ATCC  
 CC 25618, ATCC 27294, 66108, AAM358827, 16627-92, AAL68372, AAL11150,  
 CC AAL24204, AAL3308, AAL6980, L1761, TMC 306, AAL10373 and AAL23263. This  
 CC of M tuberculosis which are used in a novel method to rapidly identify strains  
 CC acid hydrazide). The method involves the use of restriction fragment  
 CC length polymorphism (RFLP) analysis to determine if a NciI-MspI  
 CC restriction site is absent in the DNA of the strain at the codon  
 CC corresponding to codon 483 of this katG gene consensus sequence. The  
 CC absence of the site indicates an INH-resistant strain.  
 QY Query Match 50.3%; Score 2041.5; DB 19; Length 740;  
 DB Best Local Similarity 34.4%; Pred. No. 3.5e-168;  
 DB Matches 405; Conservative 92; Mismatches 218; Indels 29; Gaps 13;  
 QY 28 HRAIRER---ANSKCPVM-HGNTSTGTSGKOMPGLNLDILHODRKSDDMPDPNY 83  
 DB 5 HPPTETTTGAASNGCPVGHKKYVGGGQNDMPNRLNKLVLQNPVADPACATDY 64  
 QY 84 REEVRKLDKDKVHALMTOQWMPADMGHYGGLMIRMAHSAAGTYRIADGRGGGT 143  
 DB 65 AAEVATIDVATLDEEVTWTSOPMPADYGHGGLFIRMAHAAAGTYRIADGRGAGG 124  
 QY 144 GSOREFAPNSPDWNSDKARLLWPKKKYGNKISWADLMLTLAGTYVYSGLPAGFS 203  
 DB 125 GOREFAPNSPDWNSDKARLLWPKKKYGNKISWADLMLTLAGTYVYSGLPAGFS 184  
 QY 204 FGVSDIPEKDYIWDGSEKWLAPSDERYGDVKNPMTENPLAANVOMGLIYNPEGVNGH 263  
 DB 185 FGVSDIPEKDYIWDGSEKWLAPSDERYGDVKNPMTENPLAANVOMGLIYNPEGVNGH 238  
 QY 264 PDPLTAAQVLTFPAMANDKTAALPAGHYVGVGNGNKSALADPKASDYEENGL 323  
 DB 239 PDPMANAVDIRFRMANMDVETALVVGHTFGKTHGAGPADLVGPEAPLEOMGL 298  
 QY 324 GWGNPMQGNKASNVGTIGCAWTTNPTFKDMYFDLLFGYMWELKSPAGAHWEIPD- 382  
 DB 299 GWGNPMQGNKASNVGTIGCAWTTNPTFKDMYFDLLFGYMWELKSPAGAHWEIPD- 358  
 QY 383 IKEKNKVPDASDSIRHNPIMFDAMALNVPYTPACIKCFMADPFFKTFKAKNFKLT 442  
 DB 359 AGAGTIPDPFGPGQ--RSPFMALDLSLRVDPIYERITRMLPELADEPAKATKYL 416  
 QY 443 HRDGLPKRSYIGPEVPAEDLIMODIPAGNTDYCEE---VKOKIAQSLISEMAYSTA 498  
 DB 417 HRDGLPKRSYIGPEVPAEDLIMODIPAGNTDYCEE---VKOKIAQSLISEMAYSTA 476  
 QY 499 WDSNTYRGSDMGRCANGARILAPONWOGNEPE-BLAKVLSYVEOI-----SADTG- 550  
 DB 477 WAAASFRSGDKRGANGRILOPQVGEVNDPDGLKVRTLEETQESFNSAAGNI 536  
 QY 551 -ASTADVILAGSVGTEKAAAGDYVVPFLKRGDATAEMTDSPAPLEPADGRF 609  
 DB 537 KYSPADVLVWGCNATEKAAAGNINITYPTPTDASQEDVSPVLEPAGDFRN 596

QY 610 HQKEVYVPEKMLDRAQLGTLGRTVYLLGCMRYLGNVTGGKGGVTCDEGLTND 659  
 DB 597 YLKGNEPLPAEYMLDRAQLGTLGRTVYLLGCMRYLGNVTGGKGGVTCDEGLTND 656  
 QY 670 FVNLITDMGNSWK--PVGSNAYETDRDKTGAVKATSRVDLWFGNSLSLRYAEVYQDD 727  
 DB 657 FVNLITDMGNTWPEPADGDTGCKD-GSGKVTMTSRVDLWFGNSLSLRYAEVYQDD 715  
 QY 728 NGEKVRDPAVANTVYKNAADRDV 751  
 DB 716 AQPQVQDFVAAMDKVNLDRDV 739

RESULT 8  
 AAW95399 standard; Protein: 740 AA.  
 AC AAW95399;  
 XX  
 DT 26-MAR-1999 (first entry)  
 XX  
 DE M. tuberculosis catalase peroxide (katG) gene variant.  
 XX  
 XX Catalase-peroxide: katG; mutation: isonicotinic acid hydrazide;  
 XX INH: tuberculosis; diagnosis: detection; variant.  
 XX  
 OS Mycobacterium tuberculosis.

PH Key Location/Qualifiers  
 FT /label= S137  
 FT /note= "Wild-type Ser is replaced by Thr"  
 PN W09850585-A1.  
 XX  
 PD 12-NOV-1998.  
 XX  
 PD 06-MAY-1998; 98WO-050285.  
 PF  
 PR 07-MAY-1997; 97US-0852219.  
 XX  
 PA (MAYO-) MAYO FOUNDATION.

COckerhill FR, Kline BC, Uhl JR;  
 XX WPI. 1999-070092/06.  
 DR N-PSDB; RA008010.  
 XX  
 CC Detection of Mycobacterium tuberculosis - by amplifying katG gene  
 CC and detecting specific fragment, and optionally identifying  
 CC INH-resistant strains by detecting specific mutation  
 CC  
 CC Claim 3; Page -, 83pp; English.

The invention relates to a novel method of detecting Mycobacterium tuberculosis. The method comprises amplifying the DNA in the samples to generate a detectable amount of amplified DNA comprising a catalase-peroxide (katG) DNA fragment with sequence of bases 904-1523 of the M. tuberculosis strain H37Rv MC katG gene sequence. The method optionally further comprises determining if the katG DNA fragment has a serine to threonine mutation in codon 315 (S137R mutation), indicative of an isonicotinic acid hydrazide (INH)-resistant M. tuberculosis strain. The method is especially useful to diagnose tuberculosis. This disease is a major cause of human morbidity and mortality, and conclusive diagnosis and subsequent treatment depends on identification of the etiologic agent and M. tuberculosis. INH has been used in tuberculosis treatment, but INH-resistant strains have emerged; the method allows such drug-resistant strains to be identified. The present sequence represents the amino acid sequence of the M. tuberculosis katG gene variant. The method of identification created by modifying the katG gene product given in Fig. 7.

50 Sequence 740 AA:  
 Query Match 50.3%; Score 2041.5; DB 20; Length 740;  
 Best Local Similarity 54.1%; Pred. No. 3.5e+16;  
 Matches 405; Conservative 94; Mismatches 220; Indels 29; Gaps 13;  
 QY 24 VSPREHAIERH--JMSGKCPVM-HGGTSTGTSTNKMWPGLNLDLTGQDRKSDMDP 79  
 DB 1 VPEQHPPIYETTTGAASNGCPVGHKTKPVVEGGQDMWPNRLKJLHONPAVDPMGA 60  
 QY 80 DENTREVEKLDLFDALKDOYHALMTDSOEWPMQWGGYGLMTWMHNSAGTYATDGRG 139  
 DB 61 AFDYAAZVAITDVALTRDI EYVNTTSPWPAHYGHTPLFIRMAHAGSTRIDHGRG 120  
 QY 140 GSGTGSORFAPLANSWPONVSLDKARLLWPIKKYNGKISWADLMLATVAYESMGUPA 199  
 DB 121 GAGGQHFAPLANSWPONVSLDKARLLWPIKKYNGKISWADLMLATVAYESMGUPA 180  
 QY 200 YGFSFGVDITWPEKIDWGEKEWLPASDERYGDVNPETMENPLAAYVGLIYVNEG 253  
 DB 181 EFGFGFVDMEFD-EYVWKEATWL--GOERY---SGKRDLENPLAAYVGLIYVNEG 234  
 QY 260 VNGHPDLPTAQOYLETFAAMNDEKTAALTAGHTVGNCHGNASALADPPDKASDVE 319  
 DB 235 PGNAPDMAAVDITREFFAMNMDVEYTAALVGGHTFRTGAGFADLVGPEEPAAPLE 294  
 QY 320 NOGLGNGNMOCKASNAVTSLEGATWNTPTKFGYFOLLGYWELKSPAGAHNWE 379  
 DB 295 QMGLCKSSYGTGCKAITTGIEVWVNTPTKWNPSLEILYGVEMELKSPAGAWOYT 354  
 QY 380 PID-IKKENKVPDSDPSRHNPTMDADMAIKVNPYTRAIKCEKPMADPEYFKTFKAW 438  
 DB 355 AKDQAGAGTIPDFGPG--ASPTMLATDLSURVDPIETIRTRMLHEPELADFAKAW 412  
 QY 439 FKLTRHRLGPKSRYTGPEVPARDLWODPTPAGNTDYCEEV---VKQRIASGLSISEM 494  
 DB 413 YKLTRHDMGPVARYLGLPVKPKOTLLMQDPVPVASHDLVGBAEIASLSQIRASGLTSVQL 472  
 QY 495 VSTANDSANTYRGSMDRGANAKIRLAPQNEWQGNPEP-FLAKVLSVYEQI-----SAD 548  
 DB 473 VSTANAAASSTSGDSKRGANAGRIQLAQVGHVNDPDLRRTVEETIQESFNKA 532  
 QY 549 TG---ASTIADYVLAGSVGTEKAAKAAGYDVRVPELKGSDATAEMTDADSFAPLEPAD 605  
 DB 533 PGNIKYSPFADYVLAGGCAATEKAAKAAGHNTVPTPTGRTDASQDTDVESFAVLEPKAD 592  
 QY 605 GFNNQKQKVVYVPEKMLDRAQLGTLGRTVYLLGCMRYLGNVTGGKGGVTCDEGLTND 665  
 DB 597 FVNLITDMGNSWK--PVGSNAYETDRDKTGAVKATSRVDLWFGNSLSLRYAEVYQDD 727  
 QY 666 LYNDFVNLITDMGNSWK--PVGSNAYETDRDKTGAVKATSRVDLWFGNSLSLRYAEVY 751  
 DB 653 LYNDFVNLITDMGNTWPEPADGDTGCKD-GSGKVTMTSRVDLWFGNSLSLRYAEVY 711  
 QY 728 NGEKVRDPAVANTVYKNAADRDV 751  
 DB 716 AQPQVQDFVAAMDKVNLDRDV 739  
 RESULT 9  
 AAR43670 standard; Protein: 735 AA.  
 ID AAR43670;  
 AC AAR43670;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 17-MAY-1994 (first entry)  
 XX  
 DE M. tuberculosis H37Rv catalase-peroxidase.  
 XX catalase-peroxidase: isonicotinic acid hydrazide; isoniazid; INH;  
 XX antibiotic; susceptibility; sensitive; resistant; katG.

OS	Mycobacterium tuberculosis (strain H37Rv).
XX	W09322454-AI.
PN	11-NOV-1993.
XX	30-APR-1993; 93MO-EPO1063.
PD	17-SBP-1992; 92ER-0011098.
PR	30-APR-1992; 92US-0875940.
PR	14-AUG-1992; 92US-0929206.
PR	16-APR-1993; 93FR-0004545.
XX	(ASST-) ASSISTANCE PUBLIQUE.
PA	(MEDI-) MEDICAL RES COUNCIL.
PA	(UYXB-) UNIV BERNIE.
PA	(UYVA-) UNIV CURIE PARIS VI p & M.
PI	Bodmer T, Cole S, Heym B, Honore N, Telenti A;
PI	Young D, Zhang Y;
DR	WPT: 1993-368B12/46.
DR	N-FSDB; AAQ51531.
XX	Rapid detection of antibiotic resistance in Mycobacteria - esp.
PPT	Isoniazid, rifampicin or streptomycin resistance in tuberculosis
PPT	by detecting mutation in katG, rpoB or rpsL genes
XX	Example 2; Fig 8; 97pp; English.
CC	The katG gene of Mycobacterium tuberculosis was isolated on a KpnI
CC	fragment by shotgun cloning. The sequence contains one open reading
CC	frame with high coding probability; a 735 amino acid protein is
CC	encoded by the ORF having predicted mol.wt. 80029. This protein
CC	conforms to the structure of known bacterial catalase-peroxidases
CC	which consists of two modules possibly the result of a duplication
CC	of the first module. The second module has a unique amino
CC	N-terminal sequence of about 50 amino acid residues. Mutations in
CC	the katG gene can result in loss of INH-susceptibility, i.e.
CC	the formation of antibiotic resistant Mycobacterial strains
CC	(Updated on 25-MAR-2003 to correct PN field.)
XX	Sequence 735 AA;
SQ	Query Match 46.6%; Score 1890; DB 14; Length 735;
	Best Local Similarity 51.3%; Pred. No. 5.1e-155;
	Matches 380; Conservative 95; Mismatches 240; Indels 26; Gaps 12;
OY	27 RHAIIRER--ANGSKCPWM-HGNTSGTGNKDNAMPGLMLDILHQQRDSDDPDPFN 82
XX	: :
DB	4 QHPITETTGANSCGPVGHKKYIPVEGGQNMPNPLMKVLHONPANVPAGMAAF 63
XX	: :
OY	83 YREVRKLDPDLAKXDVHALTSDSPWPADNGHYGLWLTPMANSHAGTYRIADCGGCC 142
XX	: :
DB	64 YAAEATYSRLDALTRDIEEVMTISQPWWPADYGHYGPLFIEMAHHAAGTYRIHDGGAG 123
XX	: :
OY	14 TGSRFAFLNPWDPNYLDKARRLLMPKKTKGNKISHADLTLAGTVATESGLPAYGF 202
XX	: :
DB	124 GQNRFPALNPWDNLSLRARLLMPKKTKGNKISHADLTLAGTVATESGLPAYGF 183
XX	: :
OY	203 SGVRFDLPPEKQIDVDEKELNAPSRYGVGNKPWTMENPLAAYOMGLLYINRVGVNG 262
XX	: :
DB	184 SASGVQDWETD-EVTYCKEATNL--CDDGY----SVSDLNPLAAYOMGLLYINRVPG 236
XX	: :
OY	263 HPDPLRTAQOVLFTARMANDKETAALTAGGTGTCNGCNAGSLAPDFKASVDENQG 322
XX	: :
DB	237 NPDPMAAADLRTFERMANDVETAILTGCHTFKTHGAGDAOLVGPEEPANLEDMG 296
XX	: :
OY	323 LGKNPNWGOKASNAVTSIGEGCATNTPTFKDMKYGLDYLCYNNWELKSPAGAHREID 382
XX	: :
DB	297 LGWKSSVGTGTTGKDATSIGIEVVYVNTPTFWNSFEILYGYEWELTKSPAGAWYAKD 356
XX	: :
OY	383 -IKENKPVDSAPSI RHNDIMTDADAIAKVNFTRACEKADQGVFKTKAKMYKL 441
XX	: :
DB	357 GAGACTIPDPGGFG--RSPTMLATDSLARDPYTERITRWLEHPELADEFKANYKL 414
XX	: :
OY	442 THRDJGPKSYRGVPEFAEDLIWDPIA-GNTDYCEEV--VKQIAQSGLSISEMVSTA 498
XX	: :
DB	415 IHRMDGPVARYGLPVKOTLLWDQVPAVSTTSAAKOIASLKQISRASLUYSVOLVSTA 474
XX	: :
OY	499 WDSARTYFGSDKGCANGCANITULAPQNKQNEPERLANVLSYTRQIS-----ADTGS 552
XX	: :
DB	475 WAASFSFGSKGGANGGRINLOPQGVMEVWDPSGAQHSHPEESIQESTFREGNKVS 534
XX	: :
OY	553 IADVITLVAGSVGTAKARAAGYDVRVPFLAGRCDAITAEMTDADSEAPLEPLADCFRNWOK 612
XX	: :
DB	595 FADLVWJGGCAPLEKKAAMAHNITVFPTGPDNASODPTDVSFAVLEKADGFINTYL 594
XX	: :
OY	613 KEYVYKEPEMLLDROLKMGTLGCTPMVLLGGMRVLGTNYGTGKHGVFTOCESGLINDREFY 672
XX	: :
DB	595 KGNCRSTSTSLDKANLLTISAPEMTVLVGGLRVLGANKYKLPQLGVFTAESESINDFFV 655
XX	: :
OY	673 NLTDGMSWK--PVGSNAYEIRDRTGAVKTASRVDLVFGSNSLLSTAEVYAADDNGE 730
XX	: :
DB	655 NLLDNGITWEPSADDTQTGRD--GSGKVTGTSRVDLVFGSNSELVALVEYVAPTRQA 713
XX	: :
OY	731 KPVEDFVAATKYMNADRFEDV 751
XX	: :
DB	714 KPVTGPFVAADKWMLDRFEDV 734
XX	: :
RESULT 10	
AM78363	ID AM78363 standard; Protein; 735 AA.
AC	AAW78363:
XX	20-MAR-2003 (updated)
DT	11-MAY-1999 (first entry)
DT	M. tuberculosis katG protein.
XX	katG; Catalase; peroxidase; enzyme; isonicotinic acid hydrazide;
KW	Isoniazid; INH; diagnosis; resistance.
XX	Mycobacterium tuberculosis.
OS	US5871912-A.
XX	16-FEB-1999.
PD	02-JUN-1995; 95US-0459499.
PF	02-JUN-1995; 95US-0459499.
PR	10-APR-1992; 92US-0875940.
PR	11-MAR-1



QY 598 A-----PLEPLADGFRNKKKEYV-----VKPEMLLDRAQLMGL 632  
 Db 552 SLAAVRIKSPKALPCNNSSQASATIKSKSTRKRKSCSTKPS--ADRPNDGL 610  
 QY 633 TGPFTVILGLGNRVLTGGYTKGVFTDCGLQNDFFVNLTDGNSKFPVGSNAYEIR 692  
 Db 611 SNKFARV-----GPNYRHLPHGVFTDRIGVLTNDFFVNLDMKNTGVPDTSIGYEIR 662  
 QY 693 DRKTKGAKTASRVLTGSGNSLRSYARYVADQDNGKFFVDFVFAATKYNNAADRFV 751  
 Db 663 DRKTCGEVMTATRVLTGSAVDQDNGEEFYAQQDNGKFFVDFVFAATKYNNAADRFV 721

RESULT 12  
 ABG25057  
 ID ABG25057 standard; Protein: 533 AA.  
 AC  
 AA ABG25057;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #25048.  
 XX  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 XX (HYSB-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI: 2001-639362/73.  
 XX N-PSDB; AAS89244.  
 DR  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX responsible for genetic disorders or other traits and to assess  
 XX biodiversity -  
 PS Claim 20; SEQ ID No 55416; 103pp; English.

The invention relates to isolated polynucleotide (i) and polypeptide (ii) sequences. (i) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in identifying expressed genes. (ii) is useful as gene therapy techniques for restoring normal activity of (ii) or to treat disease states involving (ii). (ii) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (i) and its binding partners are useful in medical imaging of sites expressing (ii). (i) and (ii) are useful for treating diseases and polynucleotide sequences have biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Sequence data to which this patent did not appear in the printed publication but which was submitted to the patent office in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 533 AA;  
 Query Match 25.9%; Score 1050; DB 22; Length 533;  
 Best Local Similarity 52.9%; Pred. No. 3,4e-82;  
 Matches 216; Conservative 56; Mismatches 126; Indels 8; Gaps 5;  
 QY 350 PTFDMQGYDILLFGYNWELAKKSPAGAHHPIDIKKENKPYDASDSTRHNPIMTADMA 409  
 Db 127 TQNSNFTNENFKYVMTQVTRSPAGALQAEAVD-APETIP-DPFDGSKKPKTLMVTDLT 184  
 QY 410 LKYNVTRVAKLEKADNDEYKFKTAKAFKHLHDIKDGKSRVIGSVPAEDLIHQDDIP 459  
 Db 185 LRFDFEPEKISRRLEANDFOAFNEAFAPFLJHHDGPKSRVIGPVYKPEDLLHQDP 244  
 QY 470 AGNTDYCEEVY---KOKIAQSGLSISEMVTAMDSARTYRGSDMRGANGARILAPONE 526  
 Db 245 QPIYPTQEDQIIDLKFAADSLGSLVSEIVSMASSTFRGGDKRGANGARILAPOTR 304  
 QY 527 WQGNPEPLAKVLSYVEQISNQTG-ASTADVTYVLAGSYGIEKAAKAGYDVPFLKGRG 585  
 Db 305 TGDVTRQPIRALPVLKXIKESGKASLADIIVLPLGVVEKSSASAGLSIHVFEAPGRV 364  
 QY 586 DATACMTDADSFAPLEPLADGFRNKKKEYVYVYKPEMLLDRAQLMGLTGPFTVILGQMR 645  
 Db 365 DAQQQDTAIGAFELLPDAGFRNRYRARIJVDVSTESLLIDRAQQLTLPAPFALVGMR 424  
 QY 646 VLGTFNGYGTGKHGVPTDCGGLQNDFFVNLTDGNSKFPVGSNAYEIRDRKTKGAKTAS 703  
 Db 425 VLGANFGSGKNQGVFTDRVGLSNDFFVNLDMKNTGVPDTSIGYEIRDRKTKGAKTAS 484  
 QY 704 SRVDVFGSGLLSRYAEVYAADNGKFFVDFVFAATKYNNAADRFV 751  
 Db 485 SRADLVFGSGLVLRVAREVYFASSONHEAFYKDFVFAATKYNNAADRFV 532

RESULT 13  
 ABG08823  
 ID ABG08823 standard; Protein: 273 AA.  
 AC  
 AA ABG08823;  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #8814.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 XX (HYSB-) HYSEQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 XX WPI: 2001-639362/73.  
 XX N-PSDB; AAS73010.  
 DR  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX responsible for genetic disorders or other traits and to assess  
 XX biodiversity -  
 PS Claim 20; SEQ ID No 39182; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantifying its expression, and as a food supplement. (I) and (II) are useful in medical  
 CC diagnosis, forensics, gene mapping, identification of mutations  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC of a site or population. (I) and (II) are useful for treating  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 273 AA:  
 Query Match 19.3%; Score 784.5; DB 22; Length 273;  
 Best Local Similarity 58.4%; Pred. No. 1.4e-59;  
 Matches 160; Conservative 36; Mismatches 73; Indels 5; Gaps 3;  
 OY 481 KKTAAAGSLTISBWTASDARTYRGSMDRGANGARIRAPONEWQGNPERLAKVLS 540  
 DB 1 KFTAAUGLSVLSVWMAASSTFGDGRGGANGARIALPQMDWYN--AAARALP 58  
 OY 541 VYEGISNDTG--ASTADYVLGASVGTERRAKAAGYDVRVPTLGRGDATAEKTDASFP 599  
 DB 59 VLKRIKESGKASLADIIVLAGVVGVEKAASGLSIHVPPAPGRVDAQDQDITLMPFL 118  
 OY 600 LEPDAGFRNKKRYKVPPEMLIDRAQLMCLTGPMTYLLGGHVLGTNYGTRKQVP 659  
 DB 119 LEPDAGFRNFRALDVTSTESLIDNAQQLTLPFHALVGGVMMVYAGNFGSKNVP 178  
 OY 660 TDCGOLVNDPFFVNLDMGSKFPV--GSNAVEIRDKTGAVMRTASRVLDLVFGNSLLR 717  
 DB 179 TDGVLSLNDPFFVNLDMGSKFPV--GSNAVEIRDKTGAVMRTASRVLDLVFGNSLLR 718  
 OY 718 STAEVYQADNGEKVDFVFAVMTKYMADREYV 751  
 DB 239 AVAEYASSDAHEKVFADPFAVMYKMLDRDEL 272

RESULT 14  
 ABG09297  
 ID ABG09297 standard; Protein: 1550 AA.  
 AC ABG09297;  
 AC Homo sapiens.  
 DT 13-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #9288.  
 DE Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PN W0200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-0508631.  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0643167.

XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI: 2001-639362/73.  
 XX N-PSDB: AAS73484.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX responsible for genetic disorders or other traits and to assess  
 XX biodiversity  
 XX Claim 20; SEQ ID NO 395656; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantifying a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (I) and its binding partners are useful in medical  
 CC diagnosis, forensics, gene mapping, identification of mutations  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC of a site or population. (I) and (II) are useful for treating  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1550 AA:  
 Query Match 12.8%; Score 519; DB 22; Length 1550;  
 Best Local Similarity 43.8%; Pred. No. 2.0e-35;  
 Matches 107; Conservative 34; Mismatches 66; Indels 8; Gaps 5;  
 OY 350 PTKFDMGYPDLLEGYNWELKSPAGAHHPIDIKENKPVDAOSPIRHPNIMTDAMA 409  
 DB 529 PTKMSNTFFENLEKYWQTSRPAQAIOFAYD-APEIIP-DFPDGSKRRPTMLYTDLT 586  
 OY 410 IKVNPYRACEFADPEYFKYPAKAMFKLTHDILGKRSRYIGPVPADALINDPDP 469  
 DB 587 LRFDPFEKISRRFLNDQNFARAFKULTHDNGHKSRYIGPVPADALINDPDP 646  
 OY 470 AGNTDICEEVY--KOKIAGSLISEMSTAWDSARTYRGSMDRGANGARIRAPONE 526  
 DB 647 QPNTPTQDITDLKFIADSLGSVLSVMAASASTFRRGGDKRGANGARIALMPORD 706  
 OY 527 WQNEPERLAKVLSYEDISADTG--ASTADYVLVA 560  
 DB 707 MDVN--AAARALPVLKLEKESKASLADIIVLA 739

RESULT 15  
 ABG08828  
 ID ABG08828 standard; Protein: 201 AA.  
 AC ABG08828;  
 AC Homo sapiens.  
 DT 13-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #8819.  
 DE Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 PN W0200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-0508631.  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0643167.

OS Homo sapiens.  
 XX  
 PN W0200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0645167.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 DR WPI: 2001-639352/73.  
 XX N-PSDB; AMS73015.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 XX  
 XX Claim 20: SEQ ID No 39187; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC mutations. (I) is useful in detecting antibodies against it, detecting or  
 CC quantitating a polypeptide in a sample, and as a food supplement. (II) as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and procedures for genetic disorders or other traits to assess biodiversity  
 CC and procedures for genetic disorders or other traits dependent on DNA and  
 CC amino acid sequences. AbG000010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 201 AA;  
 Query Match 9.8%; Score 398; DB 22; Length 201;  
 Best Local Similarity 41.2%; Pred. No. 3.6e-26;  
 Matches 89; Conservative 24; Mismatches 43; Indels 60; Gaps 3;  
 QY 506 RGSNMGCGANGARLARIQ---NEWGNGEPERAKVLSVEQISAGSIADIVIVLAGS 562  
 DB 21 RGSNMGCGANGARLALXQFTGDTGDTROPTRNCVYIENEL----- 62  
 QY 563 VGIKAAAGADYRVPELKGKGDATAMTDADSFAPLEADGFRNMQKKEVYVKPEEM 622  
 DB 63 -----LEPIADGFRNFRALDYSVTESL 85  
 QY 623 LIDRAQLMGLTGPENVLVGGNRVLGNTNGVTHGVFTDCRQQLTHDFVNLTDGNSWK 682  
 DB 86 LIDRAQLTLTAPENALVGGNRVLGANGDSRGVFTDVGVLGNSDFVNLDRYENRK 145  
 QY 683 PV--GSNAYEIRDKTKGAVKWTASHDVLVFGSNLL 716  
 DB 146 ATDESKELPFGCRDREGTGEVFTASRADLVFGSNVL 181

Search completed: October 7, 2003, 19:42:20  
 Job time : 64.8231 secs

\* \* \*



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

```
Run on:      October 7, 2003, 19:40:18 ; Search time 22.1175 Seconds
            (without alignments)
            1440.492 Million cell updates/sec
```

**Title:** US-09-884-889-6

Perfect score: 4058  
Sequence: 1 MNNASADDLHSSLQRCRAF.....RDFVAANTKVMNADREFDVAS 753

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: , 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

**Database :**

```
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/ptCRUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/ptCRUS_COMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Length	DB	TD	Description
1	4058	100.0	753	2	US-08-674-887A-6	Sequence 6, Appl	
2	4058	100.0	753	3	US-08-951-844-6	Sequence 6, Appl	
3	4058	100.0	753	4	US-09-412-347-6	Sequence 6, Appl	
4	2290.5	56.4	745	3	US-08-674-887A-8	Sequence 8, Appl	
5	2290.5	56.4	745	3	US-08-951-844-8	Sequence 8, Appl	
6	2290.5	56.4	745	4	US-09-412-347-8	Sequence 8, Appl	
7	2093	51.6	726	2	US-08-313-185-49	Sequence 49, Appl	
8	2093	51.6	726	2	US-08-459-499-13	Sequence 13, Appl	
9	2093	51.6	726	3	US-08-674-887A-49	Sequence 49, Appl	
10	2072.5	51.9	726	3	US-08-951-844-49	Sequence 49, Appl	
11	2044.5	50.4	740	1	US-08-418-782-2115	Sequence 11, Appl	
12	2044.5	50.4	740	1	US-08-852-219-21	Sequence 21, Appl	
13	2041.5	50.3	740	1	US-08-418-782-7	Sequence 7, Appl	
14	2041.5	50.3	740	1	US-08-228-662-7	Sequence 7, Appl	
15	2041.5	50.3	740	1	US-08-852-219-7	Sequence 7, Appl	
16	2040.5	50.3	729	2	US-08-313-185-50	Sequence 50, Appl	
17	2040.5	50.3	729	2	US-08-459-499-14	Sequence 50, Appl	
18	2040.5	50.3	729	3	US-09-082-614A-50	Sequence 50, Appl	
19	1890	46.6	735	2	US-08-313-185-48	Sequence 48, Appl	
20	1890	46.6	735	2	US-08-459-499-9	Sequence 9, Appl	
21	1890	46.6	735	2	US-08-674-887A-48	Sequence 48, Appl	
22	1885	46.5	733	2	US-08-951-844-48	Sequence 48, Appl	
23	1881	41.5	652	2	US-08-313-185-53	Sequence 53, Appl	
24	1681	41.4	652	2	US-08-459-499-17	Sequence 17, Appl	
25	1681	41.4	652	3	US-09-082-614A-53	Sequence 53, Appl	
26	1616	39.8	731	2	US-08-313-185-51	Sequence 51, Appl	
27	1616	39.8	731	2	US-08-459-499-15	Sequence 15, Appl	

## ALIGNMENTS

## RESULT 1

US-08-674-887A-6  
Sequence 6, Application US/08674887A  
Patent No. 5939300  
GENERAL INFORMATION:  
INVENTOR: Richardson, Dao E.  
APPLICANT: SAGAL, INDIAJI  
APPLICANT: Aghikari, Robert S.  
TITLE OF INVENTION: CATALASES  
NUMBER OF INVENTIONS: 8  
CORRESPONDENCE ADDRESS:  
ATTORNEY: Fish & Richardson P.C.  
STREET: 221 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FastLoad for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/674,887A  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:

Query Match 100.0%: Score 4058: DB 2: Length 753:

Best Local Similarity 100.0%; Pred. No. 0;

Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNASADLLHSSLQRCRAFLVPSRHRAIRERAMSGKCPLYMHGGNTSTCT<sup>SNKDW</sup>PE 60

```

Db      1  MNNSADLUSSLQRCRAFYVLPSPHRAIRERAMSKCPVHMGGNTSTGTSNKDWPE 60
QY      61  GLNLDLTHQODKSDPMDPDFNTRYEVRLDFAKLDVHALMDSQEMPAADHGHGGL 120
Db      61  GLNLDLTHQODKSDPMDPDFNTRYEVRLDFAKLDVHALMDSQEMPAADHGHGGL 120
QY      121  MIRMAHISAGTYRTADRGGGGTGSGQRFAPLNSPDNYSLDKARLLAPIKKTYGNKISW 180
Db      121  MIRMAHISAGTYRTADRGGGGTGSGQRFAPLNSPDNYSLDKARLLAPIKKTYGNKISW 180
QY      181  ADLMLILACTWAVESMGLPAYGFSFGVDIWEPEKDIYWGDEKEMLAPSDERYGDVKNPET 240
Db      181  ADLMLILACTWAVESMGLPAYGFSFGVDIWEPEKDIYWGDEKEMLAPSDERYGDVKNPET 240
QY      241  MENPLAAYOMGLIYVNPYGVNGHPDLRTAQOVLLETFAARMANDEKTAALTAGHVTGNC 300
Db      241  MENPLAAYOMGLIYVNPYGVNGHPDLRTAQOVLLETFAARMANDEKTAALTAGHVTGNC 300
QY      301  HGNKNSALAPPOKASDVENGLGNGPNMGKASNAVTSGETANTNPTKFDMGYFOL 360
Db      301  HGNKNSALAPPOKASDVENGLGNGPNMGKASNAVTSGETANTNPTKFDMGYFOL 360
QY      361  LFGYNMELKSPAGAHMPEIDIKKENKPVASDPSIRHNPIMTDADMAIKVNPYTAIC 420
Db      361  LFGYNMELKSPAGAHMPEIDIKKENKPVASDPSIRHNPIMTDADMAIKVNPYTAIC 420
QY      421  EKFMADPEYFKTKFAMKFLTHROLGPKSRYIGPEVPAEDLIWQDPIPAAGNTDYCEVV 480
Db      421  EKFMADPEYFKTKFAMKFLTHROLGPKSRYIGPEVPAEDLIWQDPIPAAGNTDYCEVV 480
QY      481  KOKIQAQSLGISSEMVSTANDSARTYRGSOMRGANGARIRLAPONQWNEPERLAKVLS 540
Db      481  KOKIQAQSLGISSEMVSTANDSARTYRGSOMRGANGARIRLAPONQWNEPERLAKVLS 540
QY      541  VYEOSADTGCASIDVILAGSVGTEKAAKAGTDVVRPVLKGRGDATMDADSPAPL 600
Db      541  VYEOSADTGCASIDVILAGSVGTEKAAKAGTDVVRPVLKGRGDATMDADSPAPL 600
QY      601  EPLADGFRNNOKKEYVYKPEMLDRAQLKGLTGPEMVLGGRVLYGTNYGTHKGVPT 660
Db      601  EPLADGFRNNOKKEYVYKPEMLDRAQLKGLTGPEMVLGGRVLYGTNYGTHKGVPT 660
QY      661  DCEGQLTNDPFPVNTDMGNSKPVGSNAYEIRDRKTGAVKMTASRVLDVFGSNSLLASTA 720
Db      661  DCEGQLTNDPFPVNTDMGNSKPVGSNAYEIRDRKTGAVKMTASRVLDVFGSNSLLASTA 720

```

## RESULT 2

```

US-08-951-844-6
Patent No. 6,607,480
Application US/08951844
GENERAL INFORMATION:
APPLICANT: Robertson et al.
TITLE OF INVENTION: Catalogs
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" INCH DISKETTE
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/951,844
FILING DATE:
CLASSIFICATION: 435
PUBLICATION NUMBER:
APPLICATION NUMBER: 08/674,887
FILING DATE: July 3, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Heiron
REGISTRATION NUMBER: 28,019
REFERENCE/DOCKET NUMBER: 331400-55
TELEPHONE/COMMUNICATION INFORMATION:
TELEPHONE: 201-394-1740
TELEFAX: 201-394-1744
INFORMATION FOR SEQ ID NO.: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 753 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULAR TYPE: Polypeptide
US-08-951-844-6

```

```

Query Match          100.0%: Score 4058; DB 3; Length 753;
Best Local Similarity 100.0%: Pred. NO. 0;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1  MNNSADLUSSLQRCRAFYVLPSPHRAIRERAMSKCPVHMGGNTSTGTSNKDWPE 60
Db      1  MNNSADLUSSLQRCRAFYVLPSPHRAIRERAMSKCPVHMGGNTSTGTSNKDWPE 60
QY      61  GLNLDLTHQODKSDPMDPDFNTRYEVRLDFAKLDVHALMDSQEMPAADHGHGGL 120
Db      61  GLNLDLTHQODKSDPMDPDFNTRYEVRLDFAKLDVHALMDSQEMPAADHGHGGL 120
QY      121  MIRMAHISAGTYRTADRGGGGTGSGQRFAPLNSPDNYSLDKARLLAPIKKTYGNKISW 180
Db      121  MIRMAHISAGTYRTADRGGGGTGSGQRFAPLNSPDNYSLDKARLLAPIKKTYGNKISW 180
QY      181  ADLMLILACTWAVESMGLPAYGFSFGVDIWEPEKDIYWGDEKEMLAPSDERYGDVKNPET 240
Db      181  ADLMLILACTWAVESMGLPAYGFSFGVDIWEPEKDIYWGDEKEMLAPSDERYGDVKNPET 240
QY      241  MENPLAAYOMGLIYVNPYGVNGHPDLRTAQOVLLETFAARMANDEKTAALTAGHVTGNC 300
Db      241  MENPLAAYOMGLIYVNPYGVNGHPDLRTAQOVLLETFAARMANDEKTAALTAGHVTGNC 300
QY      301  HGNKNSALAPPOKASDVENGLGNGPNMGKASNAVTSGETANTNPTKFDMGYFOL 360
Db      301  HGNKNSALAPPOKASDVENGLGNGPNMGKASNAVTSGETANTNPTKFDMGYFOL 360
QY      361  LFGYNMELKSPAGAHMPEIDIKKENKPVASDPSIRHNPIMTDADMAIKVNPYTAIC 420
Db      361  LFGYNMELKSPAGAHMPEIDIKKENKPVASDPSIRHNPIMTDADMAIKVNPYTAIC 420
QY      421  EKFMADPEYFKTKFAMKFLTHROLGPKSRYIGPEVPAEDLIWQDPIPAAGNTDYCEVV 480
Db      421  EKFMADPEYFKTKFAMKFLTHROLGPKSRYIGPEVPAEDLIWQDPIPAAGNTDYCEVV 480
QY      481  KOKIQAQSLGISSEMVSTANDSARTYRGSOMRGANGARIRLAPONQWNEPERLAKVLS 540
Db      481  KOKIQAQSLGISSEMVSTANDSARTYRGSOMRGANGARIRLAPONQWNEPERLAKVLS 540
QY      541  VYEOSADTGCASIDVILAGSVGTEKAAKAGTDVVRPVLKGRGDATMDADSPAPL 600
Db      541  VYEOSADTGCASIDVILAGSVGTEKAAKAGTDVVRPVLKGRGDATMDADSPAPL 600
QY      601  EPLADGFRNNOKKEYVYKPEMLDRAQLKGLTGPEMVLGGRVLYGTNYGTHKGVPT 660
Db      601  EPLADGFRNNOKKEYVYKPEMLDRAQLKGLTGPEMVLGGRVLYGTNYGTHKGVPT 660
QY      661  DCEGQLTNDPFPVNTDMGNSKPVGSNAYEIRDRKTGAVKMTASRVLDVFGSNSLLASTA 720
Db      661  DCEGQLTNDPFPVNTDMGNSKPVGSNAYEIRDRKTGAVKMTASRVLDVFGSNSLLASTA 720

```

Qy 721 EYTAQDNGKEFYRDFVAAMTKVNNADRFVAS 753  
 Db 721 EYTAQDNGKEFYRDFVAAMTKVNNADRFVAS 753

## RESULT 3

US-09-412-347-6  
 ; Sequence 6, Application US/09412347  
 ; Patent No. 5939300  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Robertson, Dan E.  
 ; APPLICANT: Sanyal, Indrajit  
 ; APPLICANT: Ashikari, Robert S.  
 ; TITLE OF INVENTION: CATALASES  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Suite 1400  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: US  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; OPERATING SYSTEM: Windows95  
 ; SOFTWARE: FASTSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; PRIOR APPLICATION NUMBER: US/09/412,347  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/674,887  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Halle, Ph.D., Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 09015/002001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619/678-5070  
 ; TELEFAX: 619/678-5099  
 ; INFORMATION FOR SEQUENCE 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 753 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; FRAGMENT TYPE: Internal  
 ; US-09-412-347-6

Query Match 100.0%; Score 4058; DB 4; Length 753;  
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNASADDDHSSIQQRCAFPLVSPRRRAIRERAMSGKCPVHHGGNTGTGTGSKNDWPE 60  
 Db 1 MNASADDDHSSIQQRCAFPLVSPRRRAIRERAMSGKCPVHHGGNTGTGTGSKNDWPE 60  
 Qy 61 GLNDLILQDRKSDPDMPFNRYREVRKLDKDKVDHALMDQEWMPADKMGHGYGL 120  
 Db 61 GLNDLILQDRKSDPDMPFNRYREVRKLDKDKVDHALMDQEWMPADKMGHGYGL 120  
 Qy 121 MIRMAHSGATYRIADRGGGGTGSGRPAFLASNPDSVLDKARLLWPIKKYGNKISW 180  
 Db 121 MIRMAHSGATYRIADRGGGGTGSGRPAFLASNPDSVLDKARLLWPIKKYGNKISW 180  
 Qy 181 ADMILAGTVAYESGLPAYGFSGRVDPINPEKDIYNGDEKWLAPSDERYGVNKPET 240  
 Db 181 ADMILAGTVAYESGLPAYGFSGRVDPINPEKDIYNGDEKWLAPSDERYGVNKPET 240  
 Qy 241 MENPLAANVGLIYINPEGVNGHPDLRTAQOVLETFANMAWDEKTAALTAGHHYVNC 300  
 Db 241 MENPLAANVGLIYINPEGVNGHPDLRTAQOVLETFANMAWDEKTAALTAGHHYVNC 300

Qy 301 HONGNASALADPKASDVENGSLGNGFNWCKGASNAVYSGIEGAMTMTPTKFDMGYFDL 360  
 Db 301 HONGNASALADPKASDVENGSLGNGFNWCKGASNAVYSGIEGAMTMTPTKFDMGYFDL 360  
 Qy 361 LFGYNWELKSPAGAHWEPIIDIKENKPYDASDPSIRNPINTDADMAIKVNPYTRAC 420  
 Db 361 LFGYNWELKSPAGAHWEPIIDIKENKPYDASDPSIRNPINTDADMAIKVNPYTRAC 420  
 Qy 421 EKEMADPEVFKTKFAKFWKLTTHRDGPKSYTGIVGEPVPAEDLIWDPZPAGNTDICEPV 480  
 Db 421 EKEMADPEVFKTKFAKFWKLTTHRDGPKSYTGIVGEPVPAEDLIWDPZPAGNTDICEPV 480  
 Qy 481 KOKTAQSGISLSEMYSTANDSARTYKSGDMRGANGARILAPONEWQGNPEPLKAYLS 540  
 Db 481 KOKTAQSGISLSEMYSTANDSARTYKSGDMRGANGARILAPONEWQGNPEPLKAYLS 540  
 Qy 541 VYEQISADPGASADYIVLAGSVGIEKAAKAGYDVRVFLKRGDMYAMETDAUSFAPL 600  
 Db 541 VYEQISADPGASADYIVLAGSVGIEKAAKAGYDVRVFLKRGDMYAMETDAUSFAPL 600  
 Qy 601 EPLADGFRNWKKEVYKPEMLLDRAQLMGLTGPEMTVLLGCHVLTGNTGSKHGYET 660  
 Db 601 EPLADGFRNWKKEVYKPEMLLDRAQLMGLTGPEMTVLLGCHVLTGNTGSKHGYET 660  
 Qy 661 DECBGLTNDPVMVITDMGNSMKVPVGSNAYEIRDRKTGAVKWTASRVOLVFGSNILRSYA 720  
 Db 661 DECBGLTNDPVMVITDMGNSMKVPVGSNAYEIRDRKTGAVKWTASRVOLVFGSNILRSYA 720  
 Qy 721 EYTAQDNGKEFYRDFVAAMTKVNNADRFVAS 753  
 Db 721 EYTAQDNGKEFYRDFVAAMTKVNNADRFVAS 753

## RESULT 4

US-08-674-887A-8  
 ; Sequence 8, Application US/08674887A

; Patent No. 5939300

; GENERAL INFORMATION:

; APPLICANT: Robertson, Dan E.

; APPLICANT: Sanyal, Indrajit

; APPLICANT: Ashikari, Robert S.

; TITLE OF INVENTION: CATALASES

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows95

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA: 08/674,887A

; FILING DATE: 03 JUL 1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Halle, Ph.D., Lisa A.

; REGISTRATION NUMBER: 38,347

; REFERENCE/DOCKET NUMBER: 09015/002001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 619/678-5070

; TELEFAX: 619/678-5099

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 745 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-674-867A-8

Query Match 56.4%; Score 2290.5; DB 2; Length 745;  
Best Local Similarity 59.1%; Pred. No. 1.8e-207;  
Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;

```

QY 38 GKCPVHMG---NTSTGTSNKNWPEGLNLDILHOODKSDPMDPQFNTRREVRKLDPA 94
DB 18 GKCPFTGSLKOSAGGCTKNWPNMNLNGLIRQHSSLDNDFDPDIATAEFKLDAA 77
QY 95 LKDDVHALMTDSQEMWPDHGYGSLMIRMAHISAGTYRTADGRGGGTGSGRFAPLNSW 154
DB 78 VKDDLAALMTDSQEMWPDHGYGSLMIRMAHISAGTYRTADGRGGGTGSGRFAPLNSW 154
QY 178 VKDDLAALMTDSQEMWPDHGYGSLMIRMAHISAGTYRTADGRGGGTGSGRFAPLNSW 137
QY 155 PNVSLDRAKRLNPTKKYGNKISWADLMTAGTVAESMGLPAYGFSRGVDIWEPEK 214
DB 138 PNVSLDRAKRLNPTKKYGNKISWADLMTAGTVAESMGLPAYGFSRGVDIWEPEK 197
QY 215 DYNKDEKEMAPSDERY-GDVWKPEPMFLAONMGLIYVNPPEGVHPDPLRTAQOV 273
DB 198 DYNKDEKEMAPSDERY-GDVWKPEPMFLAONMGLIYVNPPEGVHPDPLRTAQOV 251
QY 274 LETFABAHNDKTAALTAGHTVGNCHNGNNSA-LAPDKASDVENQGLGNHPNMOG 332
DB 252 RETFGNHNDEETVALLAGHTFKYTHGAADAKYVGRPAAGTEMISLGHKTYGTG 311
QY 333 KASNAVTEGATNTPTQMSNFFENLFTYENELTKSPAGTQWPKDGAGCTTPDA 392
DB 312 HADTITSGLEGANTKTPTQMSNFFENLFTYENELTKSPAGTQWPKDGAGCTTPDA 371
QY 393 SDPSIRHNPTMTDADNAKVNPTYTRAIKCEKNADPEYFKTKAKANFKLTHRLGPKSRY 452
DB 372 HDPSKSHAPMLTTDLALRNDPDEKISRRTYENDEDEADAFANAKYLTHRDNGPKVRY 431
QY 453 IGPVPAEDUINQDPTIPAGNTDYCE---EYVKQIAQSGLSISEMWTAMDSARTYRS 508
DB 432 IGPVPAEDUINQDPTIPAGNTDYCE---EYVKQIAQSGLSISEMWTAMDSARTYRS 491
QY 509 DMWGGANGARILAPQKQWENQGNPERLAKVLSYVEQISAD-----TGASADTVILA 560
DB 492 DMWGGANGARILAPQKQWENQGNPERLAKVLSYVEQISAD-----TGASADTVILA 551
QY 561 GSVGIEKAAAGYDVRVPPFLKGRGDATAMTDAOSFAPLEPLADGFRNMOKKEYVKPE 620
DB 551 GSVGIEKAAAGYDVRVPPFLKGRGDATAMTDAOSFAPLEPLADGFRNMOKKEYVKPE 620
QY 552 GCAQYENKADHGEVQVTFVGRADATQGVDETFEPAADGFRNTPKERYKSAE 611
DB 601 EXMLDRAQLMGLTGPWPMVILGSGHMLVGTNGCTKHGVTFOCEGLTNDPYNLTDMGNS 680
QY 612 EXMLDRAQLMGLTGPWPMVILGSGHMLVGTNGCTKHGVTFOCEGLTNDPYNLTDMGNS 671
QY 681 WK---PVGSNAYEIRDRKTGAVKKTASRLVFGSGLSLRSTAYEYTAQDNGEKVDFVA 738
DB 672 WKASDESQYFGEGAFRTGCVNWSGTINVLIFGNSLRLALAEVYCCADSEENFKYDFVK 731
QY 739 APTKYVNAADREF 751
DB 732 ANAYVMDLDRFL 744

```

RESULT 5  
US-08-951-844-8  
Sequence 8  
Patent No. 6074860

GENERAL INFORMATION: application US/08951844  
APPLICANT: Robertson et al.  
TITLE OF INVENTION: Catalases  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESS: CHICAGO, ILLINOIS, USA  
CITY: CHICAGO, ILLINOIS, USA  
STREET: 6 BECKER FARM ROAD

CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
COMPUTER: IBM PS/2  
COMPUTER RELEASABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/951,844  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/674,887  
FILING DATE: July 3, 1996  
ATTORNEY/AGENT INFORMATION:  
Name: Charles J. Herron  
Reference/Check Number: 28,018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 745 AMINO ACIDS  
TOPOLOGY: 1 LANKAR  
MOLECULE TYPE:  
US-08-951-844-8

Query Match 56.4%; Score 2290.5; DB 3; Length 745;  
Best Local Similarity 59.1%; Pred. No. 1.8e-207;  
Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;

```

QY 38 GKCPVHMG---NTSTGTSNKNWPEGLNLDILHOODKSDPMDPQFNTRREVRKLDPA 94
DB 18 GKCPFTGSLKOSAGGCTKNWPNMNLNGLIRQHSSLDNDFDPDIATAEFKLDAA 77
QY 95 LKDDVHALMTDSQEMWPDHGYGSLMIRMAHISAGTYRTADGRGGGTGSGRFAPLNSW 154
DB 78 VKDDLAALMTDSQEMWPDHGYGSLMIRMAHISAGTYRTADGRGGGTGSGRFAPLNSW 137
QY 155 PNVSLDRAKRLNPTKKYGNKISWADLMTAGTVAESMGLPAYGFSRGVDIWEPEK 214
DB 138 PNVSLDRAKRLNPTKKYGNKISWADLMTAGTVAESMGLPAYGFSRGVDIWEPEK 197
QY 215 DYNKDEKEMAPSDERY-GDVWKPEPMFLAONMGLIYVNPPEGVHPDPLRTAQOV 273
DB 198 DYNKDEKEMAPSDERY-GDVWKPEPMFLAONMGLIYVNPPEGVHPDPLRTAQOV 251
QY 274 LETFABAHNDKTAALTAGHTVGNCHNGNNSA-LAPDKASDVENQGLGNHPNMOG 332
DB 252 RETFGNHNDEETVALLAGHTFKYTHGAADAKYVGRPAAGTEMISLGHKTYGTG 311
QY 333 KASNAVTEGATNTPTQMSNFFENLFTYENELTKSPAGTQWPKDGAGCTTPDA 392
DB 312 HADTITSGLEGANTKTPTQMSNFFENLFTYENELTKSPAGTQWPKDGAGCTTPDA 371
QY 393 SDPSIRHNPTMTDADNAKVNPTYTRAIKCEKNADPEYFKTKAKANFKLTHRLGPKSRY 452
DB 372 HDPSKSHAPMLTTDLALRNDPDEKISRRTYENDEDEADAFANAKYLTHRDNGPKVRY 431
QY 453 IGPVPAEDUINQDPTIPAGNTDYCE---EYVKQIAQSGLSISEMWTAMDSARTYRS 508
DB 432 IGPVPAEDUINQDPTIPAGNTDYCE---EYVKQIAQSGLSISEMWTAMDSARTYRS 491
QY 509 DMWGGANGARILAPQKQWENQGNPERLAKVLSYVEQISAD-----TGASADTVILA 560
DB 492 DMWGGANGARILAPQKQWENQGNPERLAKVLSYVEQISAD-----TGASADTVILA 551
QY 561 GSVGIEKAAAGYDVRVPPFLKGRGDATAMTDAOSFAPLEPLADGFRNMOKKEYVKPE 620
DB 551 GSVGIEKAAAGYDVRVPPFLKGRGDATAMTDAOSFAPLEPLADGFRNMOKKEYVKPE 620

```

Db 552 GCAGVEKAADGAGHEVOVFPNGRADATASQTDVEAFEALEPAADGFRNVIKPEKHSVAE 611  
 Qy 621 EMLDRAQLMGLTGPMTVLLGKGRVLTNGYTKGVTDCSGQLTDFVNLTDGNS 680  
 Db 612 EMLVDRQALLSAPETALVGGKRVLTNYDSQIGVFNKPGQLSDFVNLDLWTK 671  
 Qy 681 WK--PVGSNAVETROKTAGVAKWTASRVLDVFGSNLSRYAETVAODDNGEKFVRDPA 738  
 Db 672 WRASDESUNYFGRDFTKTEYVMSGTRVLDLFGSNLSRLAALAEVTCADSEKFKVDK 731  
 Qy 739 AMTKVNAADRDV 751  
 Db 732 ANAKVMDLDRDL 744  
 RESULT 6  
 ; US-09-412-347-8  
 ; Sequence 8, Application US/09412347  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Robertson, Dan E.  
 ; APPLICANT: Sanyal, Indrajit  
 ; APPLICANT: Adhikari, Robert S.  
 ; APPLICANT: Chakrabarti, Chakraborty  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Suite 1400  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: US  
 ; FILING DATE: 08/674,887  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: Windows95  
 ; SOFTWARE: FASTSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/412,347  
 ; PRIORITY NUMBER: 08/674,887  
 ; PRIOR APPLICATION DATA:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Halle, Ph.D., Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; TELEPHONE: 619/678-5099  
 ; TELEPHONE: 619/678-5070  
 ; TELEFAX: 619/678-5099  
 ; INFORMATION FOR SEQ ID NO: 8:  
 ; LENGTH: 745 amino acids  
 ; TYPE: amino acid  
 ; MOLECULE TYPE: protein  
 ; FRAGMENT TYPE: Internal  
 ; US-09-412-347-8  
 Query Match 56.4%; Score 2290.5; Db 4; Length 745;  
 Best Local Similarity 59.1%; Pred. No. 1.8e-207;  
 Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;  
 Qy 38 GKCPVWGG--WISTSTSNKQWREGILNLTLDQQRSDNDPDNRYEYKLDLDA 94  
 Db 18 GKCPVWGG--WISTSTSNKQWREGILNLTLDQQRSDNDPDNRYEYKLDLDA 77  
 Qy 95 LKQDVALMDSQWPAQWGHYGLMTAMWHSAGTYRADRGGGGTGSGRPAFLNS 154  
 Db 78 VKQDLAALMDSQWPAQWGHYGLMTAMWHSAGTYRADRGGGGTGSGRPAFLNS 137  
 Qy 155 PNWSLDKARLLHTTKYKQNKISNADMLILATVAVESKGLPAYGTFGRVDWPRK 214

Db 138 PNANLDRKARLLWPIKQYGRKISNADMLILATVAVESKGLPAYGTFGRVDWPRK 197  
 Qy 215 DLYNGCDKARLLWPIKQYGRKISNADMLILATVAVESKGLPAYGTFGRVDWPRK 273  
 Db 198 DLYNGCDKARLLWPIKQYGRKISNADMLILATVAVESKGLPAYGTFGRVDWPRK 251  
 Qy 274 LETFARMANDKTAALTAGGRTVGNCHGNMNSA--LAPDPKASDVNOGLGNHNNQ 332  
 Db 252 RETFARMANDKTAALTAGGRTVGNCHGNMNSA--LAPDPKASDVNOGLGNHNNQ 311  
 Qy 333 KASNAVETROKTAGVAKWTASRVLDVFGSNLSRYAETVAODDNGEKFVRDPA 392  
 Db 312 HCATITISLEGATKTPQMSNPFENLEFGYEWELTASPAGAYOMKPADGAGATDPA 371  
 Qy 372 SDPSRHPNIMTDADMAIRVNPYTRAIACEFMAADPEYFKTKAFANFKLTHLDRGLPSRY 452  
 Db 372 HOPSGAPFALMTOLAIQMDPDYKISRYTENPEPFAADFAKAWTKLTHLDRGLPSRY 431  
 Qy 453 IGPVPAEDLIQMDPDYKISRYTENPEPFAADFAKAWTKLTHLDRGLPSRY 508  
 Db 432 LGPEVPOEDLIQMDPDYKISRYTENPEPFAADFAKAWTKLTHLDRGLPSRY 491  
 Qy 509 DMROGANGARIRIAPONEKQNEPPELAKYLSVYEQISAD-----TGASTADIVL 560  
 Db 492 DMROGANGARIRIAPONEKQNEPPELAKYLSVYEQISAD-----TGASTADIVL 551  
 Qy 561 GSVGTEKAAKAGYDVEVPEFLKRGDGTAEKTDADSPAPLEADGFRWOKKYEK 620  
 Db 552 GCAGVEKAADGAGHEVOVFPNGRADATASQTDVEAFEALEPAADGFRNVIKPEKHSVAE 611  
 Qy 621 EMLDRAQLMGLTGPMTVLLGKGRVLTNGYTKGVTDCSGQLTDFVNLTDGNS 680  
 Db 612 EMLVDRQALLSAPETALVGGKRVLTNYDSQIGVFNKPGQLSDFVNLDLWTK 671  
 Qy 681 WK--PVGSNAVETROKTAGVAKWTASRVLDVFGSNLSRYAETVAODDNGEKFVRDPA 738  
 Db 672 WRASDESUNYFGRDFTKTEYVMSGTRVLDLFGSNLSRLAALAEVTCADSEKFKVDK 731  
 Qy 739 AMTKVNAADRDV 751  
 Db 732 ANAKVMDLDRDL 744  
 RESULT 7  
 ; US-09-313-185-49  
 ; Sequence 49, Application US/08313185  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boate  
 ; APPLICANT: Cote, Stewart  
 ; APPLICANT: Young, Douglas  
 ; APPLICANT: Zhang, Ying  
 ; APPLICANT: Honore, Nadine  
 ; APPLICANT: Telenti, Amalio  
 ; APPLICANT: Bodmer, Thomas  
 ; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance  
 ; NUMBER OF SEQUENCES: 56  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ; STREET: 1300 I Street, N.W.  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/313,185

: FILING DATE: 12-02-1994  
 : CLASSIFICATION: 435  
 : AGENCY/AGENCY INFORMATION:  
 : NAME: HEVETS K0813  
 : REGISTRATION NUMBER: 25,146  
 : REFERENCE/DOCKET NUMBER: 02356-0068-00000  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (202) 408-4000  
 : TELEFAX: (202) 408-4400  
 : INFORMATION: SEC ID NO: 49  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 726 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: peptide  
 : US-08-313-185-49

Query Match 51.6%; Score 2093; DB 2; Length 726;  
 Best Local Similarity 53.4%; Pred No Re-189;  
 Matches 405; Conservative 103; Mismatches 206; Indels 44; Gaps 10;

QY	4	ASADLHSSILQORCAFVPLVSRRAIRERAMSGKCPVMHGG---NSTGTGSKNDWPE	60
DB	2	STSDSDHNT-----TATGKCFHQGGHDSAGAGTTTRDMFN	39
QY	61	GLNLDLHQQDQKSDPDQPNVREYVKLDFDALKDKVHALMPTDSQSWPMDNGHYGL	120
DB	40	QLRVLLNQHNSRNPUGEDFDYRKFSKLDYGLKKDLKALLTESQWMPADNGSYAGL	99
QY	121	MIRMAHISACTYRTADRGGGTGSORFAPLNSWPNDVSLDKARLLMPTKKYGNKISW	180
DB	100	FIRMAHIGACTYRSIDRGGAGGQORFAPLNSWPNDVSLDKARLLMPTKKYGNKISW	159
QY	181	ADLMTLACTVAVESKGLPAFGSPGVDTWPEKDYNGDEKWLAPSDERYGDVKNPZ	240
DB	160	ADLFLAGNVALENSGRTFGGAGRDVWEPFLDNGDEKAMLT-----RHPEA	211
QY	241	M-ENPLAAVOMGLVYNPEGVHPDPLRTAQOVLTFARMAHNDKTAALTAGHTVGN	299
DB	212	LAKNPLGATKGLVYNPEGVHPDPLRTAQOVLTFARMAHNDKTAALTAGHTVGN	271
QY	300	CHGNASALAPDPASDVNQLGNGFNHOGKASNAVTSIGSAWTPNPKFDMCYTD	359
DB	272	THGAGTFSWGPDPPEATIEQGLAWASYGSGVGADALTSLGVVWVOTPTQNSYFFE	331
QY	360	LIFGYWELKSPGAAHNEPDIIDKKNPVDASDPSRHNPIMTDMAIKVNTFYAI	419
DB	332	NLFYENPVOTRSPALQFAND-APLIP-DFPDSEKKKPLWLTDLIRDFEPEKI	389
QY	420	CEFMADPEYFGKTFARAKFPLTHRDLPKSRVIGPEVPAEDLLWDPIDFAGNDYCEV	479
DB	390	SRRFLNDPQAFNEAFARAFKLTNRDMPKSRVIGPEVPEKDLWDPIDFAGNDYCEV	449
QY	480	V---KQIAGSLISEWNTANDSATYRSQDMRGANGARILAPQNEQNGNEPERLA	536
DB	450	IIDLKALADSLGSELYSVANASATFGQDQNGKGNALAPQNGVYV---AAAY	507
QY	537	KVLSVTRQISADTG-ASTADVTLVAGSVIEKAAAGDYVRVPTLKRGDAMTDAD	595
DB	508	RALPVLKIOKESKASLADITVLVAGVYGEKASAGLSIHVPFAPRVDARQDQDIE	567
QY	568	SPAPLEPLAGFRNKKYKVVYKPEMLDRAQLMGLTGPMTVLLGNGNRVLTNGYCKT	655
DB	568	NEELLEPIALQFRTARLVLDYSTYSLSDITRAQLTUTAPENALPQGNRYLQNGNDSR	627
QY	656	HGVETDECGQLNDFVNLDMGMSWKPV---GSNAYEIRDRKTGAVKATASRVAFSGN	713
DB	628	NGVFTDRVGLVNSDFVNLDMRYENKATDESKELFEGDRDRETGEVKTASRADLVSGN	687
QY	714	SELBSYAVYADDNCKEYVSRFVAAVNTVMAHNDQYV	751
DB	688	SVLRVAVAYASSDAHEKFKVDYAAVYVNNLDRFLD	725

RESULT 8  
 US-08-439-499-13  
 : Sequence 13, Application US/08459499  
 : Patent No. 5871912  
 : GENERAL INFORMATION:  
 : APPLICANT: Heym, Beate  
 : APPLICANT: Young, Douglas B.  
 : APPLICANT: Zhang, Ying  
 : APPLICANT: Zhang, Ying  
 : TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods  
 : TITLE OF INVENTION: Detecting Mycobacterium Tuberculosis Resistant to Isoni  
 : NUMBER OF SEQUENCES: 17  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 : STREET: 1300 I Street, N.W.  
 : STATE: Washington  
 : COUNTRY: USA  
 : ZIP: 20005-3315  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : OPERATING SYSTEM: IBM PC compatible  
 : SOFTWARE: PC-DOS/MS-DOS  
 : SOURCE: Sequences released to the public under release #1.0, Version #1.3  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/459,499  
 : FILING DATE: 02-JUN-1995  
 : CLASSIFICATION: 536  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/875,940  
 : FILING DATE: 02-JUN-1992  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/929,206  
 : FILING DATE: 27-MAY-1992  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/029,655  
 : FILING DATE: 11-MAR-1993  
 : AGENT/AGENT INFORMATION:  
 : NAME: Finnegan, Henderson, Farabow, Garrett &  
 : REGISTRATION NUMBER: 25,146  
 : REFERENCE/DOCKET NUMBER: 03495-0110-03000  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 202-408-4000  
 : TELEFAX: 202-408-4400  
 : INFORMATION: SEC ID NO: 49  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 726 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: peptide  
 : US-08-459-499-13  
 : Query Match 51.6%; Score 2093; DB 2; Length 726;  
 : Best Local Similarity 53.4%; Pred No Re-189;  
 : Matches 405; Conservative 103; Mismatches 206; Indels 44; Gaps 10;

QY	4	ASADLHSSILQORCAFVPLVSRRAIRERAMSGKCPVMHGG---NSTGTGSKNDWPE	60
DB	2	STSDSDHNT-----TATGKCFHQGGHDSAGAGTTTRDMFN	39
QY	61	GLNLDLHQQDQKSDPDQPNVREYVKLDFDALKDKVHALMPTDSQSWPMDNGHYGL	120
DB	40	QLRVLLNQHNSRNPUGEDFDYRKFSKLDYGLKKDLKALLTESQWMPADNGSYAGL	99
QY	121	MIRMAHISACTYRTADRGGGTGSORFAPLNSWPNDVSLDKARLLMPTKKYGNKISW	180
DB	100	FIRMAHIGACTYRSIDRGGAGGQORFAPLNSWPNDVSLDKARLLMPTKKYGNKISW	159
QY	181	ADLMTLACTVAVESKGLPAFGSPGVDTWPEKDYNGDEKWLAPSDERYGDVKNPZ	240

```

160 ADLFLAGNALVNSGRTFGAGREDWEPDLYVNGDEKAWLTH-----RHPEA 211
161 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 M--ENPLAAVQGLIYVNPSCYNGHPDLRTAOVLESTFARMANDEKTAALTAGCHTVGN 299
242 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
212 LAARAPGATENGILYVNPSCYNGHPDLRTAOVLESTFARMANDEKTAALTAGCHTVGN 271
213 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 CHGNASALAPDKASDVNOGLGNGPNMOKGASNAVTSIGEGMNTNPTKFDMGYFD 359
301 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 TIGAGPTSNVGPDPAPAEITPDSGAGADATTSGLVYVWTFQTNWTFE 331
273 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 LLEFYVWELKSPAGAHMPEIDIKKENKPDVSGDASIRNPTMDADMAIKVNPYRAI 419
361 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
332 NFETVWOTSPAGAOFAVD--APEIIP-DPFDPSKKRPMLVDTLIRDFPEFKI 389
333 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
420 CEKFMADPEPKTKFAAMFKLTHRDLPKPSRYIGPEVPAEDLIWQDIPAGNTDYCEV 479
421 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
390 SRRELNDPOAFNEAFAMFKLTHRDLPKPSRYIGPEVPAEDLIWQDIPAGNTDYCEV 449
391 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 V---KORTIAQGLSISDMVSTANDSARTYSGDMRGANGARILAPONENQNEPERLA 536
481 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 IDLUKFAINDSGLSVSELVSNVANSASTFSGDGXGANGARILAPONENQNEPERLA 507
451 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
537 KVLVSVEQISADTG--ASIDAVTVLAGSVGEKAAAGYDVYRVPFLPAGRGDATAEMTD 595
538 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
508 SPAPLEADQFQKSNWKEVYKVEKEDLQALQGLQDPMVILGSGVILGNGVILGNGV 655
509 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
568 MEELERINDCFNRYEARLDNSTSTELLIDKAQQLTLTAPENTALVGMVILGNGVILG 713
569 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
656 HGVFTDCBGOITNDFEFTLTDGNSKVP--GSNAVEIRDRKTAGVAKTSRDLVFGSN 627
657 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
628 NGVFTDRGVGLNSDFVNLDMRYEKKATDESKELPEGRDRETEGVKFTASRADLVFGSN 687
629 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
714 SLIASVAYEYAAQDNGKEFVDFVFAANTVYKMAADREDFV 751
715 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
688 SVLRANVAYASSDAHEKFEVDFVFAANVYKMAADREDFL 725
689 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 9

```

US-09-082-614A-49
: Sequence 49, Application US/09082614A
: Patent No. 6124098
: GENERAL INFORMATION:
: APPLICANT: Reym, Beate
: INVENTOR: Cole, Stewart
: APPLICANT: Zhang, Ylong
: APPLICANT: Zhang, Ylong
: APPLICANT: Honore, Nadine
: APPLICANT: Telenti, Analo
: APPLICANT: Bodmer, Thomas
: TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance
: TITLE OF INVENTION: In Mycobacterium Tuberculosis
: PRIORITY CLAIM: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
: ADDRESSEE: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3115
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentia Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: PUBLICATION NUMBER: US/09/082, 614A
: CLASSIFICATION:
: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: US 08/313,185
: FILING DATE: 12-OCT-1994
: ANNOTATED/CLASSIFICATION:
: NAME: Myers, Kenneth
: REGISTRATION NUMBER: 25,146
: REFERENCE/DOCKET NUMBER: 02356.0068-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION CONTACT PERSON:
: SEQUENCE CHARACTERISTICS: 49:
: LENGTH: 726 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-09-082-614A-49
Query Match 51.6%; Score 2093; Db 3; Length 726;
Best Local Similarity 53.4%; Pred. Mo. 8e-189;
Matches 405; Conservative 103; Mismatches 206; Indels 44; Gaps 10;
Oy 4 ASADLHSSQQRCAFVPLVSPRRHATREARMSGCPVHMG--MTGTSTNKNMMP 60
Db 2 STSDIHT-----TRTKCFPHQGHGQAGAGTTTHMPPN 39
Oy 61 GUNLIDHLOOHRSDPMDPDRVRRVRLDLDALAKOVHVALMDSDGPNADWNGYGL 120
Db 40 QLRVDLLNHSNSNPGLGDFDYRKFKSLDYGLKKDLKALITESPMPADWNGSYAGL 99
Oy 121 MTRMAHISAGTYRIADRGSGGSGTGSORFAPLNSWPNVSLDKARLLMLIKKYGNTK 180
Db 100 FTRMAGHAGTYRSIDGQAGRGQORFAPLNSWPNVSLDKARLLMLIKKYGNTK 159
Oy 181 ADMILTAGTYRSIDGQAGRGQORFAPLNSWPNVSLDKARLLMLIKKYGNTK 240
Db 160 ADLFLAGNALVNSGRTFGAGREDWEPDLYVNGDEKAWLTH-----RHPEA 211
Oy 241 M--ENPLAAVQGLIYVNPSCYNGHPDLRTAOVLESTFARMANDEKTAALTAGCHTVGN 299
Db 212 LAARAPGATENGILYVNPSCYNGHPDLRTAOVLESTFARMANDEKTAALTAGCHTVGN 271
Oy 300 CHGNASALAPDKASDVNOGLGNGPNMOKGASNAVTSIGEGMNTNPTKFDMGYFD 359
Db 272 TIGAGPTSNVGPDPAPAEITPDSGAGADATTSGLVYVWTFQTNWTFE 331
Oy 360 LLEFYVWELKSPAGAHMPEIDIKKENKPDVSGDASIRNPTMDADMAIKVNPYRAI 419
Db 332 NFETVWOTSPAGAOFAVD--APEIIP-DPFDPSKKRPMLVDTLIRDFPEFKI 389
Oy 420 CEKFMADPEPKTKFAAMFKLTHRDLPKPSRYIGPEVPAEDLIWQDIPAGNTDYCEV 479
Db 390 SRRELNDPOAFNEAFAMFKLTHRDLPKPSRYIGPEVPAEDLIWQDIPAGNTDYCEV 449
Oy 480 V---KORTIAQGLSISDMVSTANDSARTYSGDMRGANGARILAPONENQNEPERLA 536
Db 450 IDLUKFAINDSGLSVSELVSNVANSASTFSGDGXGANGARILAPONENQNEPERLA 507
Oy 537 KVLVSVEQISADTG--ASIDAVTVLAGSVGEKAAAGYDVYRVPFLPAGRGDATAEMTD 595
Db 508 SPAPLEADQFQKSNWKEVYKVEKEDLQALQGLQDPMVILGSGVILGNGVILGNGV 655
Oy 568 MEELERINDCFNRYEARLDNSTSTELLIDKAQQLTLTAPENTALVGMVILGNGVILG 713
Db 656 HGVFTDCBGOITNDFEFTLTDGNSKVP--GSNAVEIRDRKTAGVAKTSRDLVFGSN 627
Oy 628 NGVFTDRGVGLNSDFVNLDMRYEKKATDESKELPEGRDRETEGVKFTASRADLVFGSN 687
Db 714 SLIASVAYEYAAQDNGKEFVDFVFAANTVYKMAADREDFV 751
Db 688 SVLRANVAYASSDAHEKFEVDFVFAANVYKMAADREDFL 725

```







```

: ZIP: 55402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/418,782
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Woessner, Warren D.
: REGISTRATION NUMBER: 30,440
: REFERENCE/DOCKET NUMBER: 150,141051
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-339-0331
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 740 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-418-782-7

Query Match 50.3%; Score 2041.5; DB 1; Length 740;
Best Local Similarity 54.4%; Pred. No. 6.1e-184;
Matches 405; Conservative 92; Mismatches 218; Indels 29; Gaps 13;

QY 28 HRAIRER---AMSGKCPW-HGNTSTGTSNKNMPEGLALDIHQDQKSDPDPPWY 83
DB 5 HPIITETTTGAASGCVPHVGHKPVGGGQNMNPRLNLAIVLHQNPADVPMGAAPDY 64
QY 84 REEYKLDLDAKLDKDVHMLTDSQSWPADMGHYGLMIRMAHWSAGTYRIADRGGGGT 143
DB 65 AAEVATIDVDAUTDIEEVMITTSQPMWADTYGHYGLFIRMAHAAAGTYRIHDGKGAG 124
QY 144 GSORFAPLNSPDNYSLOKARRLWPDKKKYKKNISNADLMILAGTYVAYSCLPAYGS 203
DB 125 GNGRFAPLNSPDNASLOKARRLWPVKKYKKNISNADLMILVFNAGCALESMPGKTFPG 184
QY 204 FGKVDINPEKDIYWDDEKMLAPSDERYGDVKNPETHNPLAAVQMSGLIYVNPESYNGH 263
DB 185 FQVQDNDPD-EYVWKGKATWL--QDRTY---SKRDLENPLAAVQMSGLIYVNPESYNGH 238

: ZIP: 55402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/418,782
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Woessner, Warren D.
: REGISTRATION NUMBER: 30,440
: REFERENCE/DOCKET NUMBER: 150,141051
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 612-339-0331
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 740 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-418-782-7

Query Match 50.3%; Score 2041.5; DB 1; Length 740;
Best Local Similarity 54.4%; Pred. No. 6.1e-184;
Matches 405; Conservative 92; Mismatches 218; Indels 29; Gaps 13;

QY 28 HRAIRER---AMSGKCPW-HGNTSTGTSNKNMPEGLALDIHQDQKSDPDPPWY 83
DB 5 HPIITETTTGAASGCVPHVGHKPVGGGQNMNPRLNLAIVLHQNPADVPMGAAPDY 64
QY 84 REEYKLDLDAKLDKDVHMLTDSQSWPADMGHYGLMIRMAHWSAGTYRIADRGGGGT 143
DB 65 AAEVATIDVDAUTDIEEVMITTSQPMWADTYGHYGLFIRMAHAAAGTYRIHDGKGAG 124
QY 144 GSORFAPLNSPDNYSLOKARRLWPDKKKYKKNISNADLMILAGTYVAYSCLPAYGS 203
DB 125 GNGRFAPLNSPDNASLOKARRLWPVKKYKKNISNADLMILVFNAGCALESMPGKTFPG 184
QY 204 FGKVDINPEKDIYWDDEKMLAPSDERYGDVKNPETHNPLAAVQMSGLIYVNPESYNGH 263
DB 185 FQVQDNDPD-EYVWKGKATWL--QDRTY---SKRDLENPLAAVQMSGLIYVNPESYNGH 238

```

```

Qy 264 PDPLRTAQOVLLETFARMAANDKRTAALPAGGHTVGNCHGNASALAPDPKASDVNGL 323
Db 239 PDPAANAANDIRETFARMAANDVETALLVGGHTFGKTHGAGPADLVGPPEFAALDQWGL 298
Qy 324 GGNPNMCKASNAVTSIGEGAMTNPTRKFDGMLFGLFGYNMELKSPAGAHNEPID- 382
Db 299 GKSGSYGTGDKAIDTSGLEVYNTNPKWDSLEILYGYEWELTSPAGAMOYTKDG 358
Qy 383 IKENKVDASDSTRINDITDADMAIKVNTPTRAICEKPMADPEYFKTKPAKAMFKLT 442
Db 359 AGAGTIPDPFGPG--RSPTMLATDLSLRVDPIERTIRMLHPEELAEFAKMYKLI 416
Qy 443 HRDLGPKSYIGPEVPAEDLWQDPYPAAGTDOYCEEV-----VKQIAGSLISEMSTA 498
Db 417 HRDGMVARYLGPVPTQLWQDPVPAVSHDLVGEAEIASLASQVTSQVSTGA 476
Qy 499 WDSARTYRGSMDGANGARILAPONWQNEPE-RLAKVLSYEQI-----SADTG-- 550
Db 477 WAAASSEFGSGGNGRIRLQPOGVWEVNDPDGLRKVIRTELEIESFNAAANI 536
Qy 551 -ASTADTVVLGVSIGLEKAAAGYDVVPLKRGDPAEMTDAPLADPAGFRN 609
Db 537 KYSFADVLVGGCALEKAAKAGHNITVPTPGRTDASQDQTVESFALSPADGRN 596
Qy 610 WKEEYVYKPEMLDRAQLMGLTGPEMTVLGGNGRVLTNGYGTGKGVFDCEGLIND 669
Db 597 YLKGKPLFAEYMLDKANLLISAPETVLVGLVGLYGANRYKRLPGVPTAESESLND 656
Qy 670 FVNLDGMSNWK--PVGSNAYETDRKTVGAKVTSRVDLVFGSNELRALVEYVYAGDD 727
Db 657 EFVNLDMGITWPSADDTGYQCKD--GSGKVKMTGSRVDLVFGSNELRALVEYVYAGDD 715
Qy 728 NGEKVFVDVAAMTKVNNLDRFV 751
Db 716 AQPFVQDFAAMTKVNNLDRFV 739

RESULT 15
US-08-852-219-7
GENERAL INFORMATION:
  APPLICATION US/08852219
  PATENT NO. 5922575
  APPLICANT: Cockerill, Franklin R.
  APPLICANT: Kline, Bruce C.
  APPLICANT: Uhl, James R.
  TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
  NUMBER OF SEQUENCES: 22
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Muelting, Rasach & Gebhardt, P.A.
    STREET: 119 No. 5922575th Fourth Street, Ste. 203
    CITY: Minneapolis
    STATE: MN
    COUNTRY: USA
    ZIP: 55401
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patent Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/852,219
    APPLICATION DATE: 08/05/1997
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Sandberg, Victoria A.
    REGISTRATION NUMBER: 41,287
    REFERENCE/DOCKET NUMBER: 230.00010130
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 612/505-2226
    TELEFAX: 612/505-2226
  INFORMATION FOR SEQ ID NO. 7:
    SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 740 amino acids
; TYPE: amino acid
; STRATEGY: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-852-219-7

Query Match
Best Local Similarity 50.3%; Score 2041.5; Db 2; Length 740;
Matches 405; Conservative 92; Mismatches 218; Indels 29; Gaps 13;

Qy 28 HRAIRER--AMSGKCPW-HGNTSTGTSNKMWPEGLNDLTLHQODRKSMDPDPENY 83
Db 5 HPEITETTTGAASNOCPVCHMKYPVBBGGNQDMPNRLNLKVLHQPAPADPMPGAADF 64
Qy 84 REEVRKLDPAKLCQVHALWLTDSQWPAWDMGHGTGLMTNMAHSGATRIADRGGGGT 143
Db 65 AARVATLDVALTRQLEIVYTSQWPAWDTGPTLIRAHMHAGTTRIHDRGAGG 124
Qy 144 GSGRPAPLNSPQVNSLDKARLLWPTIKKYNKLSNADMLLACTVAVESMGLPAYGS 203
Db 125 GMRPAPLNSPQVNSLDKARLLWPTIKKYNKLSNADMLLIVPAGNCALESNGFKTFGG 184
Qy 204 FGRVDLWEPEDKIDWGDREKMTAPSDERYGDKVNPETMENPLAAVOMGLITVNPBGVGH 263
Db 185 FGRVDMEDD-EYVWKEATWLU--GDER---SKRDLENPLAAVOMGLITVNPBPGNH 238
Qy 264 PDPLRTAQOVLLETFARMAANDKRTAALPAGGHTVGNCHGNASALAPDPKASDVNGL 323
Db 239 PDPAANAANDIRETFARMAANDVETALLVGGHTFGKTHGAGPADLVGPPEFAALDQWGL 298
Qy 324 GGNPNMCKASNAVTSIGEGAMTNPTRKFDGMLFGLFGYNMELKSPAGAHNEPID- 382
Db 299 GKSGSYGTGDKAIDTSGLEVYNTNPKWDSLEILYGYEWELTSPAGAMOYTKDG 358
Qy 383 IKENKVDASDSTRINDITDADMAIKVNTPTRAICEKPMADPEYFKTKPAKAMFKLT 442
Db 359 AGAGTIPDPFGPG--RSPTMLATDLSLRVDPIERTIRMLHPEELAEFAKMYKLI 416
Qy 443 HRDLGPKSYIGPEVPAEDLWQDPYPAAGTDOYCEEV-----VKQIAGSLISEMSTA 498
Db 417 HRDGMVARYLGPVPTQLWQDPVPAVSHDLVGEAEIASLASQVTSQVSTGA 476
Qy 499 WDSARTYRGSMDGANGARILAPONWQNEPE-RLAKVLSYEQI-----SADTG-- 550
Db 477 WAAASSEFGSGGNGRIRLQPOGVWEVNDPDGLRKVIRTELEIESFNAAANI 536
Qy 551 -ASTADTVVLGVSIGLEKAAAGYDVVPLKRGDPAEMTDAPLADPAGFRN 609
Db 537 KYSFADVLVGGCALEKAAKAGHNITVPTPGRTDASQDQTVESFALSPADGRN 596
Qy 610 WKEEYVYKPEMLDRAQLMGLTGPEMTVLGGNGRVLTNGYGTGKGVFDCEGLIND 669
Db 597 YLKGKPLFAEYMLDKANLLISAPETVLVGLVGLYGANRYKRLPGVPTAESESLND 656
Qy 670 FVNLDGMSNWK--PVGSNAYETDRKTVGAKVTSRVDLVFGSNELRALVEYVYAGDD 727
Db 657 EFVNLDMGITWPSADDTGYQCKD--GSGKVKMTGSRVDLVFGSNELRALVEYVYAGDD 715
Qy 728 NGEKVFVDVAAMTKVNNLDRFV 751
Db 716 AQPFVQDFAAMTKVNNLDRFV 739

Search completed: October 7, 2003, 19:47:34
Job time : 26.1175 secs

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: October 7, 2003, 19:45:48 ; Search time 48.759 Seconds  
(including alignments)  
2445.330 Million cell updates/sec

Title: US-09-884-889-6

Perfect score: 4058

Sequence: 1 MNNA5ADLLSLQRCRAF.....RDFVAAATKYNNAEDFVAS 753

Scoring table: BLOSUM62

Gapset 10.0, Gapext 0.5

Searched: 587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/pdata/1/pubpa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/pdata/1/pubpa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/pdata/1/pubpa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/pdata/1/pubpa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/pdata/1/pubpa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/pdata/1/pubpa/US08\_PUBCOMB.pep.\*  
7: /cgn2\_6/pdata/1/pubpa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/pdata/1/pubpa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/pdata/1/pubpa/US08\_PUBCOMB.pep.\*  
10: /cgn2\_6/pdata/1/pubpa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/pdata/1/pubpa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/pdata/1/pubpa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/pdata/1/pubpa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/pdata/1/pubpa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/pdata/1/pubpa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/pdata/1/pubpa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/pdata/1/pubpa/US06\_NEW\_PUB.pep.\*  
18: /cgn2\_6/pdata/1/pubpa/US06\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	4058	100.0	753	10	US-09-884-889-6
2	2290.5	56.4	745	10	US-09-884-889-8
3	272.5	6.7	291	10	US-09-870-501-1
4	272	6.7	250	14	US-10-027-559-6
5	234.5	5.8	287	12	US-10-342-224-18
6	183.5	4.1	1036	10	US-10-255-017A-14
7	123	3.1	1036	10	US-10-255-017A-14
8	123	3.0	447	15	US-10-205-032-10
9	120	3.0	357	14	US-10-027-559-12
10	117.5	2.9	4881	15	US-10-156-761-8481
11	115	2.8	3613	15	US-10-156-761-10432
12	115	2.8	4551	11	US-09-793-708-1
13	115	2.8	4551	12	US-10-201-465-2
14	115	2.8	4551	11	US-09-884-889-6
15	115	2.8	4013	10	US-09-881-289-31

16	115	2.8	4613	10	US-09-860-846-31
17	115	2.8	4613	11	US-09-988-3849-31
18	115	2.8	4613	11	US-09-988-3849-31
19	115	2.8	11877	10	US-09-860-846-3
20	115	2.8	11877	10	US-09-860-846-6
21	115	2.8	11877	11	US-09-836-821-6
22	115	2.8	12199	11	US-09-988-3848-6
23	114.5	2.8	308	16	US-10-174-693-381
24	114.5	2.8	1300	9	US-09-815-242-4903
25	113.5	2.8	951	12	US-10-375-241-0906
26	113.5	2.8	951	12	US-10-375-241-0906
27	113.5	2.8	1420	12	US-10-378-615-4
28	110.5	2.7	1100	15	US-10-081-873-90
29	107.5	2.6	26326	10	US-09-759-508B-2
30	107	2.6	704	15	US-10-156-761-12458
31	106.5	2.6	2169	10	US-09-738-626-5455
32	103.3	2.5	895	35	US-10-156-761-9246
33	103.3	2.5	895	35	US-10-156-761-9246
34	102.5	2.5	1477	14	US-10-092-880-4
35	102	2.5	358	12	US-10-289-757-113
36	102	2.5	608	15	US-10-156-761-8616
37	102	2.5	904	15	US-10-156-761-8762
38	102	2.5	929	15	US-10-156-761-11828
39	101.5	2.5	1752	12	US-10-387-388-2
40	101	2.5	610	10	US-09-738-626-4495
41	100.5	2.5	5215	10	US-09-860-846-2
42	100.5	2.5	5215	10	US-09-860-846-2
43	100.5	2.5	5215	10	US-09-860-846-2
44	100.5	2.5	5215	11	US-09-988-3848-2
45	100.5	2.5	5215	11	US-09-836-821-2

## ALIGNMENTS

### RESULT 1

US-09-884-889-6  
; Sequence 6, Application US/0988489  
; Patent No. US2007010280A1  
; GENBANK INFORMATION: CORPORA  
; APPLICANT: ROBERTSON, Dan  
; APPLICANT: ROBERTSON, Dan  
; APPLICANT: SAMVAL, Indralit  
; APPLICANT: ADHIKARI, Robert  
; TITLE OF INVENTION: CATALASES  
; FILE REFERENCE: DIVER1100-4  
; CURRENT APPLICATION NUMBER: US/09/884, 889  
; PRIOR FILING DATE: 2001-06-19/412, 347  
; PRIOR FILING DATE: 1998-10-05  
; PRIOR FILING DATE: 1998-10-05  
; PRIOR FILING DATE: 1997-10-16  
; PRIOR FILING DATE: 1996-07-03  
; PRIOR FILING DATE: 1996-07-03  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO: 6  
; LENGTH: 753  
; TYPE: PRT  
; ORGANISM: Alcaigenes (Deleya) squamarius  
US-09-884-889-6

Query Match 100.0%; Score 4058; DB 10; Length 753;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNNA5ADLLSLQRCRAFVPLVSPHRAIRERAMSKCPVHGGTSTGTSNDKDWPE 60  
Db 1 MNNA5ADLLSLQRCRAFVPLVSPHRAIRERAMSKCPVHGGTSTGTSNDKDWPE 60  
QY 61 GLNLDLHQQDKNSDPDPDNTREVRKLFDFALKDKVHALMTDSQMPADNGHYGSL 120  
Db 61 GLNLDLHQQDKNSDPDPDNTREVRKLFDFALKDKVHALMTDSQMPADNGHYGSL 120

QY	180	121	MIRMAHISAGTYIADRGGGGTGSGQFAPLNSWPNVDYLDKARLLMPLKKKYGNKISW	180
DB	180	121	MIRMAHISAGTYIADRGGGGTGSGQFAPLNSWPNVDYLDKARLLMPLKKKYGNKISW	180
QY	181	ADMLTAGTVAYESKGLDANVSGSRGVDTPEPKDITKQDEKVLAPSDERTGVNKPPT	240	
DB	181	ADMLTAGTVAYESKGLDANVSGSRGVDTPEPKDITKQDEKVLAPSDERTGVNKPPT	240	
QY	241	MEHPALAAVOMGLIYVNPENYGVNGPDLRTAQOVLETAIRMAINDKTAALTAGGHTVNC	300	
DB	241	MEHPALAAVOMGLIYVNPENYGVNGPDLRTAQOVLETAIRMAINDKTAALTAGGHTVNC	300	
QY	301	HGNNGASALAPPAKASDVENQGLGKNGPNMGKASNAVTSGLTGMATNPFEDKGYFDL	360	
DB	301	HGNNGASALAPPAKASDVENQGLGKNGPNMGKASNAVTSGLTGMATNPFEDKGYFDL	360	
QY	361	LFQYIMELKKSPAGAHMPEPDIIDIKENKVPDASDESIRNPIPTADMAIKVNPYTRATC	420	
DB	361	LFQYIMELKKSPAGAHMPEPDIIDIKENKVPDASDESIRNPIPTADMAIKVNPYTRATC	420	
QY	421	EKPMADPEYFKTKPAKAMFLTHRDGPKSRYSIGPEYPAEDLTIWODPIPAAGNTDYCEEV	480	
DB	421	EKPMADPEYFKTKPAKAMFLTHRDGPKSRYSIGPEYPAEDLTIWODPIPAAGNTDYCEEV	480	
QY	481	KQKIQAGSLISEKMWSTAMOSARTYRGSMDGGGANGARTILUAPONMGNEPERLAKVLS	540	
DB	481	KQKIQAGSLISEKMWSTAMOSARTYRGSMDGGGANGARTILUAPONMGNEPERLAKVLS	540	
QY	541	VTEQISADTSGASIDAVYILVAGSVGTEKAAKAGYGVDIRVPLFKRGDGTAMBDTADSFAPL	600	
DB	541	VTEQISADTSGASIDAVYILVAGSVGTEKAAKAGYGVDIRVPLFKRGDGTAMBDTADSFAPL	600	
QY	601	EPLADGFRNQWKEEYVVPSEMLIDRAQLGLGTPMTYILGGHNVLTGNTGKTHGVPY	660	
DB	601	EPLADGFRNQWKEEYVVPSEMLIDRAQLGLGTPMTYILGGHNVLTGNTGKTHGVPY	660	
QY	661	DCQGLQNTDFNVLTDMGNSWPKVGSNAYEIRDRKTGAQVKMTASRYDLVFGNSLLRYSYA	720	
DB	661	DCQGLQNTDFNVLTDMGNSWPKVGSNAYEIRDRKTGAQVKMTASRYDLVFGNSLLRYSYA	720	
QY	721	ETVAODNGDKERTYEDYAAKTYVNAADRFVYAS753		
DB	721	ETVAODNGDKERTYEDYAAKTYVNAADRFVYAS753		

## RESULTS

```

US-09-884-889-8
: Sequence 8: Application US/09884889
: Patent No. US20020102680A1
: GENERAL INFORMATION:
: APPLICANT: DIVERSA CORPORATION
: APPLICANT: ROBERTSON, Dan
: APPLICANT: SARHAK, Indrajit
: APPLICANT: SANYAL, Robert
: APPLICANT: TROTT, Robert
: TITLE OF INVENTION:
: FILE REFERENCE: DIVER1000-4
: CURRENT APPLICATION NUMBER: US/09/884, 889
: CURRENT FILING DATE: 2001-06-19
: PRIOR APPLICATION NUMBER: US/09/412, 347
: PRIOR FILING DATE: 1999-10-05
: PRIOR APPLICATION NUMBER: US/09/551, 844
: PRIOR FILING DATE: 1997-10-16
: PRIOR APPLICATION NUMBER: US/09/674, 887
: PRIOR FILING DATE: 1996-07-03
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn version 3.0

```

[illegible]

```

; SEQ ID NO 1
; LENGTH 291
; ORGANISM: Hordeum vulgare L. Haruna-nijo
; OS-09-870-501-1

Query Match
  6.7%: Score 272.5; DB 10; Length 291;
Best Local Similarity 27.0%; Pred. No. 4.6e-17;
Matches 93; Conservative 36; Mismatches 96; Indels 119; Gaps 10;

QY 120 LMTFMAHSGAGYRIADRGCGGCTGOSAPLNSFNDVSLQAKBELLPPKYYGNKIS 179
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 33 TMLRLAWHDAGTIVY-NFTGCGANGSIRVEEYTHGSNAGLKAIDLLEPIAKH-PKIT 90
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 180 MADLMLTAGVAYESMGLFATGFSGRVDMPEKDIYWGDEKWLAPSDERYGVNKEP 239
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 91 YADRLQAGVAVAYEGGTVEYTPGRD-----SSVCPRGR----- 129

QY 240 TMENPLAAVOMGLIYVNPBGVNSHDPDLRTAQOULETFARMANDKTAALTAGTHVTGN 299
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 -----PDAKKGAFLHRIIFYRMGLDKDIALV-SGSHSLGK 164

QY 300 CIGNSGNALAPDKASDVENOGVGNPMOGKASNAVTSIGSGAMTNTTFRDNGYF- 358
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 165 AR-----PE-----RSGFQDAMRDPLAFDNTFL 189

QY 359 DLIFGYNNELKSPAGAHWEPTDIKKENKVPDSDPSIRHNPIMTDADMAIKVNPYRA 418
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 ELKAGESGLLKP-----TDKALLDDPFERR 216

QY 419 ICEFMADPYRYFKTPAKAMFKUTHDOLGPKSRVIOPEVPAEDL 462
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 217 YVELYAKEDDFFDYAFESHKILUS--ELGFTPRSSGPASTKSDV 258

RESULT 4
US-10-027-559-6
; Sequence 6, Application US/10027559
; Publication No. US20020144307AL
; ORGANISM: Zea mays
; APPLICANT: Slimcoos, Carl
; TITLE OF INVENTION: Maize Defense-Inducible Genes and Their
; FILE REFERENCE: 035718/239835
; CURRENT APPLICATION NUMBER: US/10/027.559
; PRIOR FILING DATE: 2001-10-23
; PRIOR FILING DATE: 2001-10-23
; PRIOR FILING DATE: 2001-10-23
; PRIOR FILING DATE: 2001-10-23
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Zea mays
US-10-027-359-6

Query Match
  6.7%: Score 272; DB 14; Length 250;
Best Local Similarity 26.2%; Pred. No. 4e-17;
Matches 96; Conservative 36; Mismatches 102; Indels 132; Gaps 11;

QY 83 YHEVRLKDFDLAKVDVHALMTDSQWNPADNGHYGGLIMRMAHSGATYRIADRGCGG 142
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 12 YSEAV-----DANKRLAKLAKRS-----CAPLKLRLAHNSGTFDVS-SRTGSG 56

QY 143 TGSORFAPLNSPFDVSLDKARLLPWPKIKYGNKISNADLMILAGTVAYESMGLPAYGF 202
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 57 FOTMKHQSLEHANGACNAGLDIAVRLLEPIKEEF-PILSYADFQLAGVAVAVETGGRLEIF 115

QY 203 SFGRVDMPEKDIYWGDEKWLAPSDERYGVNKPETMENPLAAVOMGLIYVNPBGV 262
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 116 HPGREDKQPP-----PPGRULPDATKSGDHLRQVFGKQML-----S 152

QY 263 HPDELTAQOULETFARMANDKTAALTAGTHVTGNCHNGSNASALAPDKASDVENOG 322

```

```

Db 153 HQD-----IVALSGHTLRGCHKE----- 171
; LENGTH: 291
; ORGANISM: Hordeum vulgare L. Haruna-nijo
; OS-09-870-501-1

Query Match
  6.7%: Score 272.5; DB 10; Length 291;
Best Local Similarity 27.0%; Pred. No. 4.6e-17;
Matches 93; Conservative 36; Mismatches 96; Indels 119; Gaps 10;

QY 323 LGNGNPNMOGKASNAVTSIGSGAMTNTTFRDNGYF-DLFGFYNNELKSPAGAHWEPT 381
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 172 -----RSGFEGAMTNTPLVFNDSYFKELSGDEKGLPOLP----- 206

QY 382 DIKKENKVPDSDPSIRHNPIMTDADMAIKVNPYRAICEKFMADPYRYFKTPAKAMFKL 441
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 207 -----SDKALLSDPVHPLVRYKAADKAEPFDYTKENAILKL 242

QY 442 THRDLG 447
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 243 S--ELG 246

RESULT 5
US-10-342-224-18
; Sequence 18, Application US/10342224
; Publication No. US20030162294A1
; ORGANISM: Arabidopsis thaliana
; APPLICANT: Nathalie Verbruggen
; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress
; FILE REFERENCE: CNM-012US
; CURRENT APPLICATION NUMBER: US/10/342.224
; PRIOR FILING DATE: 2002-02-02
; PRIOR FILING DATE: 2002-02-02
; PRIOR FILING DATE: 2002-02-02
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 387
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-342-224-18

Query Match
  5.8%: Score 234.5; DB 12; Length 287;
Best Local Similarity 23.9%; Pred. No. 1.7e-13;
Matches 89; Conservative 45; Mismatches 106; Indels 133; Gaps 12;

QY 80 DFNTREYVRKLDLAKVDVHALMTDSQWNPADNGHYGGLIMRMAHSGATYRIADRG 139
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 7 DAYELKEITR-----ARRELSLIANK-----NCAPIMLRILAWHDAGTYD-AQSTK 51

QY 140 GGGTSGORFAPLNSPFDVSLDKARLLPWPKIKYGNKISNADLMILAGTVAYESMGLPA 199
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 52 QGNGSIRNBEETHIGANGSLKALDLCBGYKAKH-PRITADUQLAQVAVVETGGRPD 110

QY 200 YGFSGRGVDMPEKDIYWGDEKWLAPSDERYGVNKPETMENPLAAVOMGLIYVNPBG 259
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 111 IVFYPCRK-----SNVCKEGRL----- 129

QY 260 VNGHPDPLRTAQOULETFARMANDKTAALTAGTHVTGNCHNGSNASALAPDKASDV 319
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 -----PDAKQGFQLEDVYFYKGLSKDQIALV-SGCTLGRAR-----PE----- 168

QY 320 NQGLGMPNMOGKASNAVTSIGSGAMTNTTFRDNGYF--DLLFGFYNNELKSPAGAHH 377
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 169 -----RSGFQDQMTQPELNFDSYFVRELLKSGESGLLKP----- 204

QY 378 WEPTDIKKENKVPDSDPSIRHNPIMTDADMAIKVNPYRAICEKFMADPYRYFKTPAK 437
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 205 -----TUKTLEDDPEFLVELY-ADDAADFFDYAES 235

QY 438 WFKLTHRDLOGPKS 450
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 236 HKKLSELGAFNPS 248

RESULT 6
US-09-734-017A-54

```

; Sequence 54, Application US/09734017A  
; Patent No. 6,242,241  
; GENERAL INFORMATION:  
; APPLICANT: Lerchl, Jens  
; APPLICANT: Renz, Andreas  
; APPLICANT: Ehrhardt, Thomas  
; APPLICANT: Reindl, Andreas  
; APPLICANT: Cirpus, Petra  
; APPLICANT: Bischoff, Friedrich  
; APPLICANT: Freund, Norbert  
; APPLICANT: Duenwig, Elke  
; APPLICANT: Schmidt, Ralf-Michael  
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved  
; in the synthesis of amino acids, vitamins, cofactors, nucleotides and  
; TITLE OF INVENTION: the  
; FILE REFERENCE: nucleosides  
; CURRENT APPLICATION NUMBER: US/09-734,017A  
; CURRENT FILING DATE: 2000-12-12  
; PRIOR APPLICATION NUMBER: 60/171,100  
; PRIOR FILING DATE: 1999-12-16  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect  
; SEQ ID NO 176  
; LENGTH: 176  
; TYPE: PRT  
; ORGANISM: Physcomitrella patens  
US-09-734-017A-54

Query Match 4.18; Score 167.5; DB 10; Length 176;  
Best Local Similarity 29.18; Pred. No. 1.7e-07;  
Matches 53; Conservative 24; Mismatches 61; Indels 47; Gaps 6;

QY 120 LMTIRAHASAGTYRADRGGGGTGSGREAPLNSPDVNSLDKARLLWPIKKYGNKTS 179  
DB 35 ILRLAHNSGTYD-QESKTGGLGTIRGOELAHGANAGLIDAVNLLPIKEQF-PELS 92  
QY 180 WADMLLAGTAYESMGLPAYGSPGVDIWEPEKDIYWDGDEKWLAPSDERYGDVKNPE 239  
DB 93 YADTYTLAGVAVETGCTPIPHGKID-----HETCPVEGALPDATK-- 136  
QY 240 TMENPLAAVQMGILYVNEGVNGHPDPLRTAQOVLFTFARMANNDEKTAALTAGTGTVGN 299  
DB 137 -----GL-----DHLCR-----VFTKMGILTDKDVIYLSGAHTLGR 167  
QY 300 CH 301  
DB 168 CH 169

RESULT 7  
US-10-259-165-180  
; Sequence 180, Application US/10259165  
; Patent No. 6,242,241  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Xun  
; APPLICANT: Chang, Hui-song  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glaszbrook, Jane  
; APPLICANT: Katsirji, Puniyaki  
; APPLICANT: Krepis, Joel  
; APPLICANT: Moughamer, Todd  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricke, Darrell  
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING  
; IN THE MAMMARY GLAND OF THE RAT  
; CURRENT APPLICATION NUMBER: US/10/259,165  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620

; PRIOR FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: US 60/368,327  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR FILING DATE: 2001-09-26  
; NUMBER OF SEQ ID NOS: 782  
; SOFTWARE: PatentIn Ver. 3.0.4 (C) 2001 Syngenta  
; SEQ ID NO 180  
; LENGTH: 1036  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-259-165-180

Query Match 3.08; Score 123; DB 12; Length 1036;  
Best Local Similarity 18.38; Pred. No. 0.044;  
Matches 139; Conservative 99; Mismatches 274; Indels 248; Gaps 33;

QY 46 GNTSGTGNKDWNPBGIL--LDLHOODRKSDPMDFNREVRKLDFOALKKDVAHLM 103  
DB 92 GLTSTVLNSERSPSVSTQSCNNAD-----ERVEL--TLK--NCIS 132  
QY 104 TDSQWFPAD-----WGYGGLMIRMAWIS--AGTYRADRGSGGTGS 145  
DB 133 SDAQHEVTSAGSTSGSEKPTVMRLGNP---TRMGWRASDVANREKLAANR--GDVAN 186  
QY 146 ORFAPLNSPDVNSLDKARLLWPIKKYGNKISWADLITLAGTAYESMGLPAYGSPG 205  
DB 187 LRVGDK-----PIRENLA--VSTGNNMISKNDASNTKMGH--SHG 222  
QY 206 RVDIWEPEKDIYWDGDEKWLAPSDERYGDVKNPETMENPLAAV-----QNGLIYVP 257  
DB 223 --DHANNENPLPFCNQOQFLSP---RPNQNSQVRERENALIVSSFARILDQMKSNVTP 277  
QY 258 --EYVNGUP--DPLRTAQOVLFTFARMANNDEKTAALTAGTGTVMGHGNASALAPDKA 315  
DB 278 SSVQSPFPKSYLKGGVVYQGAEEIQV-----QNNRTAPNDKI 319  
QY 316 SDVEN--QCLGWNPNQKASNAVTSIGGA--MTNPTKFDGQVFLFLLGYNNELK 369  
DB 320 KRIPNIQDSSMARMDCTITFGSGGNVLEPOCGTSLRELKPKARQTHSKFEKM----- 371  
QY 370 KSPAGAAHWPEI--DIKKENKPYDASDPISRNPIWTDADMAIKVNTPTFAICEFMAADPE 428  
DB 372 -----HFFKQILDVDSHAQPSQLRLPSVFTISASNOVKYIGSYGT----- 415  
QY 429 FEKKTAFANFKLTHRDGPKERYIGFEPVPAEDLIWQDPIAGNTDTCYCEVWVKQIAQSG 488  
DB 416 --QDLSPASKLDIATDIFWTKYLDIPKESQD-----SNGSNASTIKYQKVGROG 464  
QY 489 -LSISEMAYSTANDSNTYRGSMDRGANGARILAPONKQGN-----PGALWQNSCTVREKFAEPFG 514  
DB 465 STIAVERPVHTEF--ANHRGNGNSEGYD-----BPRLAKVLSYFQISATGASTADIIVLIA 560  
QY 531 GSUPTAQRSSSGNQOVSVELRMLESWRSRPEISQGLPILSPISQAVLLFETSVTA 574  
DB 561 GSVGL-----EKANRAGQDYRVPYELKQD-----ATAEMTADGSPARLEADQERNW 610  
QY 575 FCPQIFPQVKLKLVSFVGSIQIPVLGIRODNKTPVAVNEEDTES-----GLLLGLSLQ 629  
DB 611 ORKEVYVKBPEMLDRAQL-----MGLTGPMTYVLLGNRYLVGTNY-----G 652  
QY 630 LKEEKEMAAKLISLASLETQIAVEYKFKHNSNGFSLIEDMDVLASNDLSGASACALGG 689  
QY 653 CTKHGVFTDCEQQLTWDFEVLNLTDMGNSKPKVGSNAVEIR 692  
DB 690 ASLSGPTID-----TSEANIKIKVDNDALRVK 716

RESULT 8  
US-10-205-032-10  
; Sequence 10, Application US/10205032



```

1 Publication No. US20020144307A1
2
3 GENERAL INFORMATION:
4
5 APPLICANT: Simmons, Carl
6
7 TITLE OF INVENTION: Male Defense-Inducible Genes and Their
8
9 TITLE OF INVENTOR: Use
10
11 FILE REFERENCE: 035718/239835
12
13 CURRENT APPLICATION NUMBER: US/10/027,559
14
15 PRIORITY CLAIMING DATE: 2000/04/23
16
17 PRIOR APPLICATION NUMBER: 60/243,120
18
19 PRIOR FILING DATE: 2000/10/25
20
21 NUMBER OF SEQ ID NOS: 13
22
23 SOFTWARE: PASTS80 for Windows Version 4.0
24
25 SEQ ID NO 12
26
27 LENGTH: 357
28
29 TYPE: PRT
30
31 ORGANISM: Dan mays
32
33 US-10-027-559-12
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1
```

; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; CURRENT FILING DATE: 2002-05-23  
 ; PRIORITY DATE: 2002-05-23  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 8481  
 ; LENGTH: 4881  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-8481

Query Match 2.84; Score 117.5; Db 15; Length 4881;

Best Local Similarity 20.84; Pred No 1.7; Mismatches 282; Indels 205; Gaps 36;

Matches 147; Conservative 72; Mismatches 282; Indels 205; Gaps 36;

Db 58 NPEGLDILHOODRSDPDPFYREVRKIDEDALKKDVIHALMTDSQWMPADNGHY 117

Db 3603 WPAAL-----PB---NDNRHLDLFTAFDIH-----RYWDTISAGHP 3637

QY 118 GGLIMBIMHNSAGTYRAGDGGGCGSORFAPLNSPDVSLDKARLLWPKKKYGNK 177

Db 3638 GDL-----SAAGLTAGHPLLGS---AVALAESOELL-----FTGR 3670

QY 178 ISMADLIMLAGTVBSMLGPATGFSRGRVDINPEKDIYWGDEKWLAPSDERYGVNK 237

Db 3671 LSKTHPMLADIAHFGVLLPGTAL-----ELAVRAGDEVD-----CGTVEE 3713

QY 238 PETHNPLAIVOMKGLIYN-----PEYVGHDPDLTAQVLETFARAKNDKTAALTA 292

Db 3714 -LTLRPTLVLPVPOGSGVLIQLSVGAPOGHOTPEEPER-----RTFALVAREDDGLSSSA 3766

QY 293 GG-RTVGCNCHGNASALA-----POPKASDVE-----NQGLGWNPNMQ 331

Db 3767 ANTCGTCTHAGTVLGTARPAEHWQBPFPADAMVDLDCHWYQIAGAGLGTG-PVFG 3825

QY 332 GK-----ASNATVSGEG-----AWTNPNTFDMGYFDLLGYNWELAKSPAG 374

Db 3826 GLREWVRGRGDEVAVTLPESTEGQAADARYALHALLDALHPVLRHESGD---AAAD 3882

QY 375 AHWEPTDIKKENKPYDASDPSTIRH-----NPIMTDMAIKVNP---TYRAI 419

Db 3883 GHEKLP--FSTGTVYAGSGASTLVLRVTFDDMGLLATDASGRIVISAGSLAFRPV 3940

QY 420 CERFM--ADPEYFKKTFAPAKFKLTHRD-----LGPKSR-----451

Db 3941 SAQIQQAARTGYHDFLEIRNPL-HLPTTPTARTADMALGFGAARTAVLERNGASWA 3999

QY 452 YIGPEVPADLIWDQPTPAGNTDYCEVYKQIAQSLGISBMYSTANDSARTYRGDMR 511

Db 4000 YPDPAALAEALGAPACHVYISCEPDGASAPTSALITSLDTPASGAPADSTALA 4059

QY 512 GGANGARIR-LAPONENOCNEPERAK---VLSVBOISADTASADVILVAGSVEIK 567

Db 4060 DATRQATRVYVALLQEWAD--ERLAACRIALLTHGCVTATPDPEVSD-LAHAAVWGLVR 4116

QY 568 AKKAGYDVRVPLKRGDASAMWDSAFAPLEAUG-----FRW-----OKTEYV 616

Db 4117 SVQTEPND---REL-----LATDSTDASRNALPLAGEFOIALRGAVIRPRTVVP 4167

QY 617 VPEEMLLDRAQLMGITGEMTVIL-GGMVILG---TYNGCTKHG 658

Db 4168 KQOPQSTTDAD---WDPENATVLTGGTGVGLRVLAHRLAHATAGV 4208

RESULT 11

US-10-156-761-10432

; Sequence 10432; Application US/0156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: HORIKAWA, JUYI

; APPLICANT: HORIKAWA, YOSHIT

; APPLICANT: SHIHA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

RESULT 12

US-09-793-708-1

; Sequence 1; Application US/09793708

; Publication No. US20030104597A1

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETHLACH, Melanie C.

; APPLICANT: BETHLACH, Gary C.

; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; CURRENT FILING DATE: 2002-05-23  
 ; PRIORITY DATE: 2002-05-23  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 8482  
 ; LENGTH: 3613  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-10432

Query Match 2.84; Score 115; Db 15; Length 3613;

Best Local Similarity 22.14; Pred No 1.8;

Matches 136; Conservative 66; Mismatches 199; Indels 214; Gaps 36;

Db 126 NBSAGTYRAGDGGGCGSORFAPLNSPDVSLDKARLLWPKKKYGNKISNADMI 185

Db 1187 WNPAGYVLTGGTGGIG-----AHLARM---LAKEAGERLL--LVSRGSOAEG-----1230

QY 186 LAGTVAYESMLGPATGFSRGRVDINPEKDIYWGDEKWLAPSD-ERYGDVKNPTEMP 244

Db 1231 -AAELATLGLOR-ETTFACVD-----SORLGHVYIAGTAPHP 1270

QY 245 LAAYOMGLIYNPNVGNHGPDLRTA-----QVLETFARMAMN-DEKTAALT-----291

Db 1271 L7AV-----FITAGSYGAB-LATATPEHYETVLSAKTGRANLDAITAEZLEAFVLF 1323

QY 292 AGCHTVCHGNCHGNASALAPDKASDVENGNGNMGKASNAVTSIGEGATTNPT 351

Db 352 KEDGCVFLLGYNWELKSPAGAHWEPI-DIKKENKPYDASDPISRHNPIMTDADMI 410

Db 1373 AAD-----GTEDQLARRGVAMEPALAKALQALEQDETAL---TVTDNWA- 1417

QY 411 KUNVTY-----RAICEKPMADPE-----YFKKTFP-----435

Db 1418 RFTPGYTIARRPLIEDIPEVARSLSSEFAPDDGSDTAQTLAGLTAPQHDRLLEL 1477

QY 436 ---KANFKLTH-----ROLAPKS-----RYGPEVPADLIWDQPTP 469

Db 1478 VNSGTAAYVTHRTDITTAGRPDLGFSLTAMELRNRLNATLGLRPLA-TLVDHPPT 1536

QY 470 AGNDTYCEVYKQIAQSLGISBMYSTANDSARTYRGDMR--GGANGARIRLAPONEW 527

Db 1537 ORLAGHILHEKLPDSATETALPV---LRTTDDDPITVGNMACRFPGVVG-----PEDLM 1587

QY 528 GGNPEERLAKVLSVQISADTASDVIVLAGSVGEKAAAGYDVRVPLKRGDA 587

Db 1588 ---TAKMTDADS---FAPLEPLADGRFMWQKEVYVPEML-----LDRAQI- 629

QY 588 ---TAKMTDADS---FAPLEPLADGRFMWQKEVYVPEML-----LDRAQI- 629

Db 1622 RQGAFLACADFDAAFGISPRELA-----NDPQORLLTSEKLEARNQPD 1669

QY 630 -HGLTGPEMTVLGGS 643

Db 1670 PVGLRSGHTGVVGG 1684

```

; APPLICANT: MCDANIEL, Robert
; APPLICANT: MCDANIEL, Robert
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30622002121
; CURRENT APPLICATION NUMBER: US 09/793,709
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 08/134,990
; PRIOR FILING DATE: 1999-03-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-793-708-1

Query Match      2.8%; Score 115; DB 11; Length 4551;
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 138; Conservative 82; Mismatches 232; Indels 204; Gaps 35;

Qy 226 APSDERYGVNKEPTMENPLAAYOMGLY-----VNPESGVNCH-----263
Db 624 APTLERY-DVQVPVF-----AVNLSLARVQHGVTPQAVYGHSGEIAAAYVAGALS 677

Qy 264 APTLERY-DVQVPVF-----AVNLSLARVQHGVTPQAVYGHSGEIAAAYVAGALS 677
Db 678 DDAARVVTLSKSIHAHLAKGKGLSLALSEDVLERLAGDGLSVAAVNGFTATVWSD 737

Qy 295 -----HYVNGCHGNASALAPDPKASD-----VENO-----GLOWNPNMOGKASNA 337
Db 738 PVQTEELARCEADGVARVPIVDYASHISROVEIIESIELAEVLASGLSPQAPRV 791

Qy 338 VTSIGIGBAMTNTPTKFMGYDGLLPGYNW--ELKKGPAAGAHWEPIDIKKENKPVDA 395
Db 732 FSTLEGAMITEVP-LOGGY-----WTRNLHRVGFAPAVETL-----ATDE 832

Qy 396 STRH-----NPLMTDADNAIKVNPYATICEKPMADPEYFKTKPAK--FKLTHDL 446
Db 833 GFTHFVENSANPVL--NALPGTVTGLATLRDNGQDRLVASLAENANGLAVNSPL 889

Qy 447 GPKRSYIGEPVPA-----EDLIWDQPT-----PAGNTDYCEEVVKQIAQ--SGLSISEMVS 497
Db 890 LFSATGHISDLPTAFATETRMHGETEALAPAGEPAVOPVLTETAEPAELDRDQLRY 949

Qy 498 ANDSA-----NYR-----GSDMRGGANGA--RIRLAPONENQ 529
Db 950 ILDKVRAOTAOVLGYATGATGQIEVDRTFREAGCTSLTCVDLNRINAAFGVMA 1009

Qy 530 NEPERLAK--VLSYVEIGSADTASDVIVLAGSVIEKAAKAAAGDVVRP-----579
Db 1010 PTFEALAQGLLVVHGGAANAPAGAPAAVGAAG--DEPVAIVGMACRLPGVVASPD 1067

Qy 580 -----FLKGRGDATAMT-----DADSAPLEPLANGFRNNKQKEY-----VKP 619
Db 1068 LARLVAGGDAISEPFGDRGNDVGLYHPDPEHPTGYVROGGFTENVAGFDAAFG 1127

Qy 620 EEMK-LDRAQLMGLTGPEMTVLLAGMRVLGTNYGTHKGVPTDCBQGLTNDFFVNTDM 678
Db 1128 REALANDPOORLLLETSEWVEDAG--IDPTSLGRGVGVT---GAMTHYGSLSR 1182

Qy 679 NSKFPVGSNAEYTOR-----KGNVWMT-----ASRDLVFGSLSRSLRSYEA 724
Db 1183 E-----GLDCLTGTNTASWSSGNYITGLSGPALTYTAC--SSSLVALHVAQ 1222

```

```

RESULT 13
US-10-201-365-2
; Sequence 2, Application US/10201365
; Publication NO. US20030148469A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BISHOP, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: COMBINATORIAL POLYKETIDE LIBRARIES PRODUCED USING A MODULAR
; TITLE OF INVENTION: PKS GENE CLUSTER AS SCAFFOLD
; FILE REFERENCE: 306622002103
; CURRENT APPLICATION NUMBER: US 10/201,365
; PRIOR FILING DATE: 2000-07-22
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-10-201-365-2

```

```

Query Match      2.8%; Score 115; DB 12; Length 4551;
Best Local Similarity 21.0%; Pred. No. 2.6;
Matches 138; Conservative 82; Mismatches 232; Indels 204; Gaps 35;

Qy 226 APSDERYGVNKEPTMENPLAAYOMGLY-----VNPESGVNCH-----263
Db 624 APTLERY-DVQVPVF-----AVNLSLARVQHGVTPQAVYGHSGEIAAAYVAGALS 677

Qy 264 APTLERY-DVQVPVF-----AVNLSLARVQHGVTPQAVYGHSGEIAAAYVAGALS 677
Db 678 DDAARVVTLSKSIHAHLAKGKGLSLALSEDVLERLAGDGLSVAAVNGFTATVWSD 737

Qy 295 -----HYVNGCHGNASALAPDPKASD-----VENO-----GLOWNPNMOGKASNA 337
Db 738 PVQTEELARCEADGVARVPIVDYASHISROVEIIESIELAEVLASGLSPQAPRV 791

Qy 338 VTSIGIGBAMTNTPTKFMGYDGLLPGYNW--ELKKGPAAGAHWEPIDIKKENKPVDA 395
Db 732 FSTLEGAMITEVP-LOGGY-----WTRNLHRVGFAPAVETL-----ATDE 832

Qy 396 STRH-----NPLMTDADNAIKVNPYATICEKPMADPEYFKTKPAK--FKLTHDL 446
Db 833 GFTHFVENSANPVL--NALPGTVTGLATLRDNGQDRLVASLAENANGLAVNSPL 889

Qy 447 GPKRSYIGEPVPA-----EDLIWDQPT-----PAGNTDYCEEVVKQIAQ--SGLSISEMVS 497
Db 890 LFSATGHISDLPTAFATETRMHGETEALAPAGEPAVOPVLTETAEPAELDRDQLRY 949

Qy 498 ANDSA-----RTYR-----GSDMRGGANGA--RIRLAPONENQ 529
Db 950 ILDKVRAOTAOVLGYATGATGQIEVDRTFREAGCTSLTCVDLNRINAAFGVMA 1009

Qy 530 NEPERLAK--VLSYVEIGSADTASDVIVLAGSVIEKAAKAAAGDVVRP-----579
Db 1010 PTFEALAQGLLVVHGGAANAPAGAPAAVGAAG--DEPVAIVGMACRLPGVVASPD 1067

Qy 580 -----FLKGRGDATAMT-----DADSAPLEPLANGFRNNKQKEY-----VKP 619
Db 1068 LARLVAGGDAISEPFGDRGNDVGLYHPDPEHPTGYVROGGFTENVAGFDAAFG 1127

Qy 620 EEMK-LDRAQLMGLTGPEMTVLLAGMRVLGTNYGTHKGVPTDCBQGLTNDFFVNTDM 678
Db 1128 REALANDPOORLLLETSEWVEDAG--IDPTSLGRGVGVT---GAMTHYGSLSR 1182

```

QY 679 NSMKPGSNAYEDR-----KTGAKMT-----ASRDVLFGSGNSLLASAEVYA 724  
 Db 1183 E-----GLDGYLLTGNTASVSGRVSVTGLGEGPALTVDTAC-SSSLVALHIAVQA 1232

RESULT 14  
 US-0-160-539-1  
 : Sequence 31, Application US/10160539  
 : Patent No. US20030162262A1  
 : GENERAL INFORMATION:  
 : APPLICANT: ASHLEY, Gary  
 : APPLICANT: BETLACH, Melanie C.  
 : APPLICANT: BETLACH, Mary C.  
 : APPLICANT: MCDANIEL, Robert  
 : TITLE: COPI INVENTION OF A RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE  
 : FILE REFERENCE: 300622002120  
 : CURRENT FILING DATE: 2002-05-29  
 : PRIOR APPLICATION NUMBER: US/10/160,539  
 : PRIOR FILING DATE: 2002-05-29  
 : PRIOR FILING DATE: 2000-09-07  
 : PRIOR APPLICATION NUMBER: 09/320,878  
 : PRIOR FILING DATE: 1998-05-27  
 : PRIOR APPLICATION NUMBER: 09/141,908  
 : PRIOR FILING DATE: 1998-06-28  
 : NUMBER OF SEQ ID NOS: 34  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 1  
 : LENGTH: 4551  
 : TYPE: PRT  
 : ORGANISM: Streptomyces venezuelae  
 US-10-160-539-1

Query Match 2.88; Score 115; DB 12; Length 4551;  
 Best Local Similarity 21.08; Pred. No. 2.6;  
 Matches 138; Conservative 82; Mismatches 232; Indels 204; Gaps 35;

QY 226 ARSDERYGDNVPTMENPLAIVOMGLY-----VNPGEVNGH----- 263  
 Db 624 APTLERV-DVVOQVTF-----AVMSLARVMOHHGVTPOAVVGHSGEIAAAYVAGALS 677

QY 264 PDPLR-----TAQVILETFA-----RHANDEKTAALTAGG 294  
 Db 678 DDAARVYTLRSKSTAAHLAGKMGSLAALSEDAVLERLAGFDGLSLVAANGPTATVYVSD 737

QY 295 -----HTVGNCHGNKASALAPPKASD-----VENO-----GLGNGNPMCKKASNA 337  
 Db 738 PVQTEELARACADGVRARVIPDYVASHRSQVEIIESELAELAGLSLSPQAPRV-----P 791

QY 338 VTSIGEGAWTTPTKFDGMGYFDLLGYNW-----ELKSPAGAHWEPTIDIKKENKPVDSADP 395  
 Db 792 FFSTLEGANITEPV-LDGGY-----NYRLNRHHVGFAPAVETL-----ATDE 832

QY 396 STRH-----NPTMDADMAIKVNTPTALCEKFMADPEVFKTKFAKAW-----FKLTHRDL 446  
 Db 833 GFTHFVSAHPVLT-----MALPQVTGLATLRDNGGDRLVSLAEAWANGLAYDMSPL 889

QY 447 GPKSRVIGPEVPA-----EDLIWDDPI-----PAGNTDYCEEVVKQIAQ-SGLSISEMYST 497  
 Db 890 LPSATGHSDLPYATFATERHNLGIEALAPAGEPAVOPVLRTEAAPAEILDREDOQJRV 949

QY 498 ANDSA-----RTVR-----GSDMRGANGA-RRLAPONEMWG 529  
 Db 950 TLUDKVPQAQVLYATGCGIYEDVTRFACQTSILTCVGLNRNRINAAGVMAPIFNDF 1009

QY 530 NEPERLAK-VLSVYEQISADTCASIAADVIVLAGSVGEKAAKAAGYDVRVP----- 579  
 Db 1010 PTFEALADQLLVVHGEAANAPGAPAPVAAAGV--DEPVAIVGMACRLPGVVASPED 1067

QY 580 -----FLKRGDAETMT-----DAUSFAPLEPLADGRFMQKKEYV-----VKP 619  
 Db 1068 LWRVLVAGGDAISEFPDQGRGNDVEGLYHDPHEPGTSYVQGGFIENVAGFDAAPFGTSP 1127

QY 620 EEML-LDRRAQLMGLTGPENTVLLGGMHRVLTNYGGTKHGVTDCBQGLTNDFFVNLDMG 678  
 Db 1128 REANLADPOQRALLTSTNENEDAG--IDPTSLRGKGVNPT---GAPHTYGPSLRDGG 1182

QY 679 NSMKPGSNAYEDR-----KTGAKMT-----ASRDVLFGSGNSLLASAEVYA 724  
 Db 1183 E-----GLDGYLLTGNTASVSGRVSVTGLGEGPALTVDTAC-SSSLVALHIAVQA 1232

RESULT 15  
 US-09-861-289-31  
 : Sequence 31, Application US/09861289  
 : Patent No. US2002010897A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Sherman, D.H.  
 : APPLICANT: Liu, H.  
 : APPLICANT: Xue, Y.  
 : TITLE: INVENTION OF A DNA encoding methymycin and pikromycin  
 : FILE REFERENCE: 600.43BU1  
 : CURRENT APPLICATION NUMBER: US/09/861,289  
 : PRIOR FILING DATE: 2001-05-18  
 : PRIOR APPLICATION NUMBER: 09/105,537  
 : NUMBER OF SEQ ID NOS: 43  
 : SOFTWARE: FastSeq for Windows Version 3.0  
 : SEQ ID NO 31  
 : LENGTH: 4613  
 : TYPE: PRT  
 : ORGANISM: Streptomyces venezuelae  
 US-09-861-289-31

Query Match 2.88; Score 115; DB 10; Length 4613;  
 Best Local Similarity 21.08; Pred. No. 2.6;  
 Matches 138; Conservative 82; Mismatches 232; Indels 204; Gaps 35;

QY 226 ARSDERYGDNVPTMENPLAIVOMGLY-----VNPGEVNGH----- 263  
 Db 686 APTLERV-DVVOQVTF-----AVMSLARVMOHHGVTPOAVVGHSGEIAAAYVAGALS 739

QY 264 PDPLR-----TAQVILETFA-----RHANDEKTAALTAGG 294  
 Db 740 DDAARVYTLRSKSTAAHLAGKMGSLAALSEDAVLERLAGFDGLSLVAANGPTATVYVSD 799

QY 295 -----HTVGNCHGNKASALAPPKASD-----VENO-----GLGNGNPMCKKASNA 337  
 Db 800 PVQTEELARACADGVRARVIPDYVASHRSQVEIIESELAELAGLSLSPQAPRV-----P 853

QY 338 VTSIGEGAWTTPTKFDGMGYFDLLGYNW-----ELKSPAGAHWEPTIDIKKENKPVDSADP 395  
 Db 854 FFSTLEGANITEPV-LDGGY-----NYRLNRHHVGFAPAVETL-----ATDE 894

QY 396 STRH-----NPTMDADMAIKVNTPTALCEKFMADPEVFKTKFAKAW-----FKLTHRDL 446  
 Db 895 GFTHFVSAHPVLT-----MALPQVTGLATLRDNGGDRLVSLAEAWANGLAYDMSPL 951

QY 447 GPKSRVIGPEVPA-----EDLIWDDPI-----PAGNTDYCEEVVKQIAQ-SGLSISEMYST 497  
 Db 952 LPSATGHSDLPYATFATERHNLGIEALAPAGEPAVOPVLRTEAAPAEILDREDOQJRV 1011

QY 498 ANDSA-----RTVR-----GSDMRGANGA-RRLAPONEMWG 529  
 Db 1012 TLUDKVPQAQVLYATGCGIYEDVTRFACQTSILTCVGLNRNRINAAGVMAPIFNDF 1071

QY 530 NEPERLAK-VLSVYEQISADTCASIAADVIVLAGSVGEKAAKAAGYDVRVP----- 579  
 Db 1072 PTFEALADQLLVVHGEAANAPGAPAPVAAAGV--DEPVAIVGMACRLPGVVASPED 1129

QY 580 -----FLKRGDAETMT-----DAUSFAPLEPLADGRFMQKKEYV-----VKP 619  
 Db 1130 LWRVLVAGGDAISEFPDQGRGNDVEGLYHDPHEPGTSYVQGGFIENVAGFDAAPFGTSP 1189

QY 620 EEML-LDRRAQLMGLTGPENTVLLGGMHRVLTNYGGTKHGVTDCBQGLTNDFFVNLDMG 678

Db 1190 REALLANDPOORLLETSWEAVEDAG--IDPISLRGRQGVFT---GAMTHEYGPSLRGG 1244  
QY 679 NSMKPVGSNAYEIRD---KTCAYWT-----ASRVDLVFGSNILRSYARYA 724  
Db 1245 E-----GLDGYLLAGNTASVMSGRVSYTIGLEGPALTYDTAC-SSSLVALHLAQA 1294

Search completed: October 7, 2003, 19:59:24  
Job time : 54.759 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: October 7, 2003, 19:39:02, Search time 29.6575 Seconds  
1441707  
2441707 Million cell updates/sec

Title: US-09-884-889-6

Perfect score: 4058

Sequence: 1 MNNSADLLHSLLQRCRAF.....RDFVAATKVMNDRFDVAS 753

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_76:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2319	71-9	724	B82186	catalase/peroxidase
2	2820	69-5	723	A93227	catalase/imported
3	2820	69-5	723	A93227	catalase/imported
4	2820	69-5	723	A93227	catalase/imported
5	2731	67-3	720	S71130	catalase (EC 1.11.1.1)
6	2514	62-0	735	B83763	catalase BH0906 [1]
7	2489	61-3	735	J90520	catalase (EC 1.11.1.1)
8	2489	61-3	731	A73874	catalase (EC 1.11.1.1)
9	2489	61-3	731	A73874	catalase (EC 1.11.1.1)
10	2489	61-3	731	A73874	catalase (EC 1.11.1.1)
11	2489	61-3	731	A73874	catalase (EC 1.11.1.1)
12	2489	61-3	732	A73877	catalase (EC 1.11.1.1)
13	2489	61-3	733	A73882	catalase (EC 1.11.1.1)
14	2279	56-2	781	F92584	catalase/peroxidase
15	2255	55-6	727	F95422	catalase (EC 1.11.1.1)
16	2255	55-6	727	F95422	catalase (EC 1.11.1.1)
17	2207	54-4	740	A75091	catalase (EC 1.11.1.1)
18	2162.5	53-3	737	A87676	hydroperoxidase HP
19	2096	51-7	726	G91237	catalase, hydropor
20	2096	51-7	726	G91237	catalase, hydropor
21	2096	51-6	726	C9E036	catalase (hydropor
22	2083	51-3	726	A10936	catalase (EC 1.11.1.1)
23	2063.5	50-9	746	A10936	catalase (EC 1.11.1.1)
24	2051.5	50-6	727	C9E036	catalase (EC 1.11.1.1)
25	2042.5	50-3	740	A40662	catalase (EC 1.11.1.1)
26	1998	49-2	731	T44846	catalase (EC 1.11.1.1)
27	1930	47-6	720	T44562	catalase (EC 1.11.1.1)
28	1922.5	47-4	726	A73874	catalase (EC 1.11.1.1)
29	1301.5	46-9	741	A59529	catalase (EC 1.11.1.1)

#### RESULT 1

B82186

catalase/peroxidase VCI560 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
C:Species: Vibrio cholerae  
C:Accession: B82186  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers  
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A:Reference number: AB2035; MOID:20406833; PMID:10952301

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-724 <ORI>

A:Cross-references: GB:A5004233; GB:A5003852; NID:99656055; PIDN:AAF94714.1; GSPDB:GN  
A:Experimental source: serogroup O1, strain N16961, biotype El Tor

C:GeneID:1560

A:Map position: 1

C:Superfamily: catalase HPI

C:Keywords: heme; iron; metalloprotein

F:99/Active site: His (distal axial ligand) #status predicted

F:267/Binding site: heme iron (His) (proximal axial ligand) #status predicted

F:318/Active site: Trp (tryptophyl radical intermediate) #status predicted

Query Match 71.9%; Score 2019; DB 2; Length 724;

Best Local Similarity 74.1%; Pos. No. 5,26-194;

Matches 530; Conservative 75; Mismatches 110; Indels 0; Gaps 0;

QY	37	SGKCPVHGNTSTGTSGNKMWPGLNDLHODKSDPMDPQVREVRKLDALK	96
DB	9	SQCQPVHGGUTSASNSNMWPKALNDLHODKTPKLGADFNREELKLDYALK	68
QY	97	KDVALMTDSQWPAQDSGTGGLTMAWNSGTTRYADRGSGGSGSORPAPLSWPD	156
DB	69	RDALKALNTSQWEPADWGUTGGLTMAWNSAGTRYADRGSGGSGSORPAPLSWPD	128
QY	157	NVSLDKARLLWPKIKKYKNISWADLMLAGTVAYESGLPAYGFSGRVDIWEPEKDI	216
DB	129	NANLDKARLLWPKIKKYKNISWADLMLAGTVAYESGLPAYGFSGRVDIWEPEKDI	188
QY	217	YKQDEKSLWADSGRYGDNVKNPQKTSMAVQNGLYTQNPQVNGHDPLETAQVLET	276
DB	189	YKQDEKSLWADSGRYGDNVKNPQKTSMAVQNGLYTQNPQVNGHDPLETAQVLET	248
QY	277	FARMANDEKTAALTAGHTYVNGCHGNSALAPDPKASOVENQGLGNPNMQKASN	336
DB	249	FARMANDEETVALTAGHTYVNGCHGNSALAPDPKASOVENQGLGNPNMQKASN	308
QY	337	AVTSGIEGANTPTKFDNGYFDLLFGYVWELKSPAGAHWEPTDIKKENKPYDASDP	396

```

Db 309 TVTSGTEGATNTPTKMGVDFYLLSTLWQLTSPAGAQWEPVNIKEDEKPVDEPS 368
Qy 397 IHNPIMTDADMAKVNPTKCEKPMADPEYKTKAFAMFKLTHROGLPKASYTGE 456
Db 398 TVTSGTEGATNTPTKMGVDFYLLSTLWQLTSPAGAQWEPVNIKEDEKPVDEPS 368
Db 399 TVTSGTEGATNTPTKMGVDFYLLSTLWQLTSPAGAQWEPVNIKEDEKPVDEPS 368
Db 399 TVTSGTEGATNTPTKMGVDFYLLSTLWQLTSPAGAQWEPVNIKEDEKPVDEPS 368
Qy 429 VPAEDLIMQDPIPAAGTDCYEEVYKOKIAGSGLSISEMVSATMSARTYRGSOMRGANG 516
Db 430 VPAEDLIMQDPIPAAGTDCYEEVYKOKIAGSGLSISEMVSATMSARTYRGSOMRGANG 516
Db 431 VPAEDLIMQDPIPAAGTDCYEEVYKOKIAGSGLSISEMVSATMSARTYRGSOMRGANG 516
Qy 517 ARTRIAPOKQWEGNEPARLVKVLAVLEKTAESGSIADTIVLAGNYTGEQAKAAQVNY 548
Db 489 ARTRIAPOKQWEGNEPARLVKVLAVLEKTAESGSIADTIVLAGNYTGEQAKAAQVNY 548
Qy 577 RYFPLAGAGDATAENTDADSFAPLEADQFNNQKKEVYVYKPEMLIDRAQLMGLTGE 636
Db 549 RYFPLAGAGDATAENTDADSFAPLEADQFNNQKKEVYVYKPEMLIDRAQLMGLTGE 636
Qy 637 MYVLGAGNRVLGNTGTHGVTGDCGSGLTNDFFVNLTDGMSNKPVGSNAYEIRORTK 696
Db 609 MYVLGAGNRVLGNTGTHGVTGDCGSGLTNDFFVNLTDGMSNKPVGSNAYEIRORTK 696
Qy 697 GAVMTASRVDLVFGSNLSLSAEVYAAQDNGEKEFVRDFAAMTKYMNADREDV 751
Db 669 GAVMTASRVDLVFGSNLSLSAEVYAAQDNGEKEFVRDFAAMTKYMNADREDV 751

```

## RESULT 2

```

AF3127
Catalase [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Accession: AF3127
C:Date: 11-Jan-2002
C:Revision: 11-Jan-2002
C:Text_change: 18-Nov-2002
A:Authors: J. C. Kaul, R. J. Monks, D. Chen, L. Wood, G. E. Chen, Y. X. Woo, J.
erage, G. Gillet, M. Grant, C. Guenther, D. Kutyavin, T. Levy, R. J. Li, M. McClellan,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Reference number: AB2577; NUID:21608550; PMID:11743193
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Accession: AF3127
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-723 <KOR>
A:GC-content: 50.8%; GC: 368
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: katA
A:Map position: linear chromosome
C:Superfamily: catalase HPI

```

```

Query Match: Similarity 69.5%; Score 2820; DB 2; Length 723;
Best local: Similarity 11.04; P: 3.7e-157;
Matches: 508; Conservative 86; Mismatches 121; Indels 0; Gaps 0;
Qy 37 SCGCKPVMGNGSTGTSKNDWPEGLINDLHQDQRKSDMPDPDNYREYKRLDFALK 96
Db 8 AGCKPVMGNGTASGVYEWPNALINDLHQDHTKTNPLGTSFNREALKTLDEALK 67
Qy 97 KDVALMTDSQWPNADKNGHVGGLIMAHMSAGTTRTAGGGGTSQGFAPLNSHPD 156
Db 68 ADLSALMTDSQWPNADKNGHVGGLIMAHMSAGTTRTAGGGGTSQGFAPLNSHPD 127
Qy 157 NVYSLDKARLLPIMPKKYGKISWADLMILAGTVAYESNGLPAGYSGFGRVDIWEPEKDI 216
Db 128 NVYSLDKARLLPIMPKKYGKISWADLMILAGTVAYESNGLPAGYSGFGRVDIWEPEKDI 187
Qy 217 YGDEKEMAPSDRYGVDSYKRPETLENPLANVQGLITVNPBGNGKSDPLATAQKRET 276
Db 188 YGDEKEMAPSDRYGVDSYKRPETLENPLANVQGLITVNPBGNGKSDPLATAQKRET 247

```

```

Qy 277 FARMAWDEKTAALTAGGHTVGNCGNAGNALADPDKASDVENQGLNGHNPNGOGRASN 336
Db 248 FARMAWDEKTAALTAGGHTVGNCGNAGNALADPDKASDVENQGLNGHNPNGOGRASN 307
Qy 337 AVTSGTEGATNTPTKMGVDFYLLSTLWQLTSPAGAQWEPVNIKEDEKPVDEPS 396
Db 398 AVTSGTEGATNTPTKMGVDFYLLSTLWQLTSPAGAQWEPVNIKEDEKPVDEPS 367
Qy 429 VPAEDLIMQDPIPAAGTDCYEEVYKOKIAGSGLSISEMVSATMSARTYRGSOMRGANG 516
Db 430 VPAEDLIMQDPIPAAGTDCYEEVYKOKIAGSGLSISEMVSATMSARTYRGSOMRGANG 516
Db 431 VPAEDLIMQDPIPAAGTDCYEEVYKOKIAGSGLSISEMVSATMSARTYRGSOMRGANG 516
Qy 517 ARTRIAPOKQWEGNEPARLVKVLAVLEKTAESGSIADTIVLAGNYTGEQAKAAQVNY 548
Db 489 ARTRIAPOKQWEGNEPARLVKVLAVLEKTAESGSIADTIVLAGNYTGEQAKAAQVNY 548
Qy 577 RYFPLAGAGDATAENTDADSFAPLEADQFNNQKKEVYVYKPEMLIDRAQLMGLTGE 636
Db 549 RYFPLAGAGDATAENTDADSFAPLEADQFNNQKKEVYVYKPEMLIDRAQLMGLTGE 636
Qy 637 MYVLGAGNRVLGNTGTHGVTGDCGSGLTNDFFVNLTDGMSNKPVGSNAYEIRORTK 696
Db 609 MYVLGAGNRVLGNTGTHGVTGDCGSGLTNDFFVNLTDGMSNKPVGSNAYEIRORTK 696
Qy 697 GAVMTASRVDLVFGSNLSLSAEVYAAQDNGEKEFVRDFAAMTKYMNADREDV 751
Db 669 GAVMTASRVDLVFGSNLSLSAEVYAAQDNGEKEFVRDFAAMTKYMNADREDV 751

```

## RESULT 3

```

C98160
Catalase (AB033631) [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001
C:Revision: 22-Oct-2001
C:Text_change: 18-Nov-2002
A:Authors: P. Hinkley, G. J. Gattung, S. J. Miller, N. Blanchard, M. J. Qurollo, B. J. Goldm
science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; NUID:21608551; PMID:11743194
A:Accession: C98160
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-723 <KOR>
A:GC-content: 50.8%; GC: 368
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: AGR_L_481
A:Map position: linear chromosome
C:Superfamily: catalase HPI

```

```

Query Match: Similarity 69.5%; Score 2820; DB 2; Length 731;
Best local: Similarity 11.04; P: 3.7e-157;
Matches: 508; Conservative 86; Mismatches 121; Indels 0; Gaps 0;
Qy 37 SCGCKPVMGNGSTGTSKNDWPEGLINDLHQDQRKSDMPDPDNYREYKRLDFALK 96
Db 16 AGCKPVMGNGTASGVYEWPNALINDLHQDHTKTNPLGTSFNREALKTLDEALK 75
Qy 97 KDVALMTDSQWPNADKNGHVGGLIMAHMSAGTTRTAGGGGTSQGFAPLNSHPD 156
Db 68 ADLSALMTDSQWPNADKNGHVGGLIMAHMSAGTTRTAGGGGTSQGFAPLNSHPD 135
Qy 157 NVYSLDKARLLPIMPKKYGKISWADLMILAGTVAYESNGLPAGYSGFGRVDIWEPEKDI 216
Db 136 NVYSLDKARLLPIMPKKYGKISWADLMILAGTVAYESNGLPAGYSGFGRVDIWEPEKDI 195
Qy 217 YGDEKEMAPSDRYGVDSYKRPETLENPLANVQGLITVNPBGNGKSDPLATAQKRET 276
Db 196 YGDEKEMAPSDRYGVDSYKRPETLENPLANVQGLITVNPBGNGKSDPLATAQKRET 255

```







Query Match	61.3%	Score	2489	3B	4	Length	751
Best Local Similarity	64.2%	Pred.	No	2E	164		
Matches	468	Conservative	85	Mismatches	160	Indels	16
Gaps	7						
QY	32	REARAGSGCVWHEGCVT	---	SYGTSENKDWHPGGLNLLHQDDRKSPMDPQNTNREVR	88		
Db	4	QNRNAAQCP	-FHGSYVTNOSSNNTT	QNRNAPQNLNLS	LAQDRKTPHDEFTNREYQ	62	
QY	89	KLDFALDKDVIALLKTSQENPDAWGHYTGSLIRIMAHMSASCTYRTADRGSGGTGQSRF	1488				
Db	63	KLUDYALKDLKRLKTSQENPDAWGHYTGSLIRIMAHMSASCTYRTADRGSGGTGQSRF	1488				
QY	149	APLNKSPDNYVSLDKARLLPPTKKYCNKISKWAQMLTACTVAESKGLPAGSFQGV	2088				
Db	123	APLNKSPDNYVSLDKARLLPPTKKYCNKISKWAQMLTACTVAESKGLPAGSFQGV	2088				
QY	209	TWPEKDTYWGDEKELAPSDERY	-GVYNKPTMENPLAAVQGLIYNTP	EGVYNGHPDL	2678		
Db	183	WHPEDYVYTESKEMLA	-SEN	SGSYD	---	RELENPLAAVQGLIYNTP	EGVYNGHPDL
QY	268	TGAQGVLTFFGSEAKNDEKTAUAGGTHYCGNCSNAPADPGPSVENCQDGN	3376				
Db	237	AAARDITRTPRMKNQDEETVALIAGGTHYCGKAGAGSATHYWGDEPAAP	TEAQGLGWLIS	2968			
QY	328	PMQGNASNAVTSIGEGAWTNTPTAFQNGYTFDLFCYAWELAKSPAGAHHPET	IDIKKN	3878			
Db	297	SYGKNGSDTISIGLGMATPTQND	SYDMLGSDTISIGLGMATPTQND	3568			
QY	388	RPVDASDSIRINPTITADMAIKVNTYVTRAI	CFRMAADPEYFCKTFAKAFMLTRHDLG	4478			
Db	357	LAPDADPSKVPYVPMKQNTDIALUREDPETEKIRARFRHONFSEABAFARFMLTRHDMG	4168				
QY	448	PKSKYIGDEYVNDLWQDPTFAGNCTYCE	---	EYVTKQ	TAQGLSGLIS	SENKYSTWDAWAT	5048
Db	417	PKTRYLGPVPKEDPTWQDPI	PPVDEY	TEELTEELKATLNSGLISV	SELVKTWASAT	4768	
QY	505	YRSGDNIRGGANGARI	RIAPQENQGNPNPERLAKVI	SVYEQISAD	---	TGASITADVYLAAGS	5628
Db	477	FNSMRDRGGANGARI	RIAPQKQWENPNPERLAKVISYVED	IRELPKRYSTADYLAAGS	5368		
QY	563	VGTEAKAANGYDVYRPLFGKQDAWAMTADSPALPLADCFNPNQKKEYVYKPEKM	6228				

Db 537 AAVEKARDGADVKKVFFFGDNTAOEOTDVESSFAVLEPFAOGFRNTQOEYSVPPEEL 596  
 QY 623 LLDRQAQLMGLTGPMTVILLCGRHVLGNTYGGTGHGVSTDCSGLTJDRFPNLTDMGNSWK 682  
 Db 597 LVYKAQLLGLTAPMTVILVGLGLRGVLTANROLPHGVFTDRIGVLTNDFVNLDMNYEW 656  
 QY 683 PVGSNATYEIDRKTGAVTASRDVLVFGSNLSLARSYAETVAQDNGEKGVDFVAAWTK 742  
 Db 657 PTDGSGIYEIDRKTGEVMTATRVLDLFGSNSILARSYAETVAQDNGEKGVDFVAAWTK 716  
 QY 743 VVNAADRFV 751  
 Db 717 VVNAADRFDL 725

RESULT 9  
 T43874  
 C:Accession: T43874  
 C:Species: synthetic  
 A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*  
 C:Date: 24-Mar-2000 *sequence\_revision* 24-Mar-2000 *text\_change* 23-Mar-2001  
 A:Reference number: 224644; PMID:95116998; PMID:9920270  
 A:Accession: T43874  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-751 <MAT>  
 A:Cross-references: EMBL:AB020079; NID:94140065; PID:BAA36991.1; PID:94140066  
 C:Comment: For the wild-type sequence, see PIR:J50520.  
 C:Gene: cat  
 C:Keywords: oxidoreductase

Query Match 61.3%; Score 2489; DB 4; Length 751;  
 Best Local Similarity 64.2%; Pred. No. 3, 2e-164;  
 Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

QY 37 REAAMSKCPVYHGGVT---STGTSKNDWPEGLNLDILHQDRKSDMDPDFFNYREVR 88  
 Db 4 QNRQNAACP-FHGVSTVHQSRRNTKNDWPEGLNLDILHQDRKTHDEEFYAEQF 62  
 QY 89 KLDFALKVDHALMTDSQWPMADMGHYGGLTSMWAHSAGTYRITADRGGGTGSORF 148  
 Db 63 KLDFALKVDHALMTDSQWPMADMGHYGGLTSMWAHSAGTYRITADRGGGTGSORF 122  
 QY 149 APUNSWPDNLDKARLLWPKIKKKYGNKISWADLTAGTVAYESMGLPAYGFSEGRVD 208  
 Db 123 APUNSWPDNLDKARLLWPKIKKKYGNKISWADLTAGTVAYESMGLPAYGFSEGRVD 182  
 QY 209 IWPEKDIYWDGKEWLAAPSERY-GDVKPMTMENPLAAYOMGLIYVPEGVNGHPDL 387  
 Db 123 APUNSWPDNLDKARLLWPKIKKKYGNKISWADLTAGTVAYESMGLPAYGFSEGRVD 182  
 QY 209 IWPEKDIYWDGKEWLAAPSERY-GDVKPMTMENPLAAYOMGLIYVPEGVNGHPDL 387  
 Db 183 VHPEDVYWSKEWLA---SERVSD-----RELENPLAAYOMGLIYVPEGVNGHPDL 236  
 QY 268 RTAQVLETFARMANDEKTAALTAGTHTVNGCHGNSALAPDPKASDVENGGLGN 327  
 Db 123 APUNSWPDNLDKARLLWPKIKKKYGNKISWADLTAGTVAYESMGLPAYGFSEGRVD 208  
 QY 237 AAARDIETFRFARMONDEEVALTAGHTGFKAHAGPATVGPPEEAPTEAQLGLWIS 296  
 QY 328 PNMCKASNAVITSGIEGANTPTKFDGMYDLPLFGYNMELKSPAGAHNHPEDIDKKEN 387  
 Db 328 PNMCKASNAVITSGIEGANTPTKFDGMYDLPLFGYNMELKSPAGAHNHPEDIDKKEN 387  
 QY 297 STYKGSQSDITSGIEGANTPTQWDYSYFDMLEFGYDNMLTKSPAGAHNHPEDIDKKEN 356  
 Db 328 PNMCKASNAVITSGIEGANTPTKFDGMYDLPLFGYNMELKSPAGAHNHPEDIDKKEN 387  
 QY 388 KPVDSAPDSIRHNPIMTDMAIKAVNPTYRACEKFMADPEYKTKAKAWFKLTHROGL 447  
 Db 388 KPVDSAPDSIRHNPIMTDMAIKAVNPTYRACEKFMADPEYKTKAKAWFKLTHROGL 447  
 QY 417 PKRYLZGPPEKDFWIDPIFSDVYELTAEETELKILNSGLTSYELVYIAMSAST 476  
 Db 417 PKRYLZGPPEKDFWIDPIFSDVYELTAEETELKILNSGLTSYELVYIAMSAST 476  
 QY 505 YRGSMDGGANGARILAPQWQNEGNEPPELAKVLSVYEOISAD--TGASINDVYLAGS 562

Db 477 FRNSDKRGANGARILAPQWQNEPPELAKVLSVYEOISAD--TGASINDVYLAGS 536  
 QY 563 VGTSEKAKAGDYVYFPLVKGKGDATKEMTADSPADLEPLAOGFRNKQKEYVVKPEM 622  
 Db 537 AAVEKARDGADVKKVFFFGDNTAOEOTDVESSFAVLEPFAOGFRNTQOEYSVPPEEL 596  
 QY 623 LLDRQAQLMGLTGPMTVILLCGRHVLGNTYGGTGHGVSTDCSGLTJDRFPNLTDMGNSWK 682  
 Db 597 LVYKAQLLGLTAPMTVILVGLGLRGVLTANROLPHGVFTDRIGVLTNDFVNLDMNYEW 656  
 QY 683 PVGSNATYEIDRKTGAVTASRDVLVFGSNLSLARSYAETVAQDNGEKGVDFVAAWTK 742  
 Db 657 PTDGSGIYEIDRKTGEVMTATRVLDLFGSNSILARSYAETVAQDNGEKGVDFVAAWTK 716  
 QY 743 VVNAADRFV 751  
 Db 717 VVNAADRFDL 725

RESULT 10  
 T43878  
 C:Accession: T43878  
 C:Species: synthetic  
 A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*  
 C:Date: 24-Mar-2000 *sequence\_revision* 24-Mar-2000 *text\_change* 23-Mar-2001  
 A:Reference number: 224644; PMID:95116998; PMID:9920270  
 A:Accession: T43878  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-751 <MAT>  
 A:Cross-references: EMBL:AB020090; NID:94140087; PID:BAA37002.1; PID:94140088  
 C:Comment: For the wild-type sequence, see PIR:J50520.  
 C:Gene: cat  
 C:Keywords: oxidoreductase

Query Match 61.3%; Score 2489; DB 4; Length 751;  
 Best Local Similarity 64.2%; Pred. No. 3, 2e-164;  
 Matches 468; Conservative 85; Mismatches 160; Indels 16; Gaps 7;

QY 37 REAAMSKCPVYHGGVT---STGTSKNDWPEGLNLDILHQDRKSDMDPDFFNYREVR 88  
 Db 4 QNRQNAACP-FHGVSTVHQSRRNTKNDWPEGLNLDILHQDRKTHDEEFYAEQF 62  
 QY 89 KLDFALKVDHALMTDSQWPMADMGHYGGLTSMWAHSAGTYRITADRGGGTGSORF 148  
 Db 63 KLDFALKVDHALMTDSQWPMADMGHYGGLTSMWAHSAGTYRITADRGGGTGSORF 122  
 QY 149 APUNSWPDNLDKARLLWPKIKKKYGNKISWADLTAGTVAYESMGLPAYGFSEGRVD 208  
 Db 123 APUNSWPDNLDKARLLWPKIKKKYGNKISWADLTAGTVAYESMGLPAYGFSEGRVD 182  
 QY 209 IWPEKDIYWDGKEWLAAPSERY-GDVKPMTMENPLAAYOMGLIYVPEGVNGHPDL 387  
 Db 123 APUNSWPDNLDKARLLWPKIKKKYGNKISWADLTAGTVAYESMGLPAYGFSEGRVD 182  
 QY 209 IWPEKDIYWDGKEWLAAPSERY-GDVKPMTMENPLAAYOMGLIYVPEGVNGHPDL 387  
 Db 183 VHPEDVYWSKEWLA---SERVSD-----RELENPLAAYOMGLIYVPEGVNGHPDL 236  
 QY 268 RTAQVLETFARMANDEKTAALTAGTHTVNGCHGNSALAPDPKASDVENGGLGN 327  
 Db 123 APUNSWPDNLDKARLLWPKIKKKYGNKISWADLTAGTVAYESMGLPAYGFSEGRVD 208  
 QY 237 AAARDIETFRFARMONDEEVALTAGHTGFKAHAGPATVGPPEEAPTEAQLGLWIS 296  
 QY 328 PNMCKASNAVITSGIEGANTPTKFDGMYDLPLFGYNMELKSPAGAHNHPEDIDKKEN 387  
 Db 328 PNMCKASNAVITSGIEGANTPTKFDGMYDLPLFGYNMELKSPAGAHNHPEDIDKKEN 387  
 QY 297 STYKGSQSDITSGIEGANTPTQWDYSYFDMLEFGYDNMLTKSPAGAHNHPEDIDKKEN 356  
 Db 328 PNMCKASNAVITSGIEGANTPTKFDGMYDLPLFGYNMELKSPAGAHNHPEDIDKKEN 387  
 QY 388 KPVDSAPDSIRHNPIMTDMAIKAVNPTYRACEKFMADPEYKTKAKAWFKLTHROGL 447  
 Db 388 KPVDSAPDSIRHNPIMTDMAIKAVNPTYRACEKFMADPEYKTKAKAWFKLTHROGL 447  
 QY 417 PKRYLZGPPEKDFWIDPIFSDVYELTAEETELKILNSGLTSYELVYIAMSAST 476  
 Db 417 PKRYLZGPPEKDFWIDPIFSDVYELTAEETELKILNSGLTSYELVYIAMSAST 476  
 QY 505 YRGSMDGGANGARILAPQWQNEGNEPPELAKVLSVYEOISAD--TGASINDVYLAGS 562

	Query Match	61 3a:	Score	2489:	D9 4:	Length	751:
	Best Local	Similarity	64.24:	Pred.	3.2e-164:		
	Matches	Conservative	85:	Mismatches	160:	Indels	Gaps
QY	32	REARMSGCKVYVGHGNT---	STGTSINKDWWPPIGLNIDT	HOODRKSDPMDPFFNYREYR	88		
QY	33						
QY	34						
QY	35						
QY	36						
QY	37						
QY	38						
QY	39						
QY	40						
QY	41						
QY	42						
QY	43						
QY	44						
QY	45						
QY	46						
QY	47						
QY	48						
QY	49						
QY	50						
QY	51						
QY	52						
QY	53						
QY	54						
QY	55						
QY	56						
QY	57						
QY	58						
QY	59						
QY	60						
QY	61						
QY	62						
QY	63						
QY	64						
QY	65						
QY	66						
QY	67						
QY	68						
QY	69						
QY	70						
QY	71						
QY	72						
QY	73						
QY	74						
QY	75						
QY	76						
QY	77						
QY	78						
QY	79						
QY	80						
QY	81						
QY	82						
QY	83						
QY	84						
QY	85						
QY	86						
QY	87						
QY	88						
QY	89						
QY	90						
QY	91						
QY	92						
QY	93						
QY	94						
QY	95						
QY	96						
QY	97						
QY	98						
QY	99						
QY	100						

	Query Match	61.3%	Score 2489;	DB 4;	Length 751;	
	Local Similarity	94.4%	Ref. 24;	Acc. 92.7%;	Indels 16;	Caps 7;
	Matches 468;	Conservative.	85;	Mismatches 160;		
QY	32	SERAMSGKCVNMGHGT	--STGTSNNDWPEGLMLDI	LHQDDKSDPDPENYEVR	88	
		: : : : : :	: : : : : :	: : : : : :	: : : : : :	
Db	4	QNRQNAQCQ-PSGYSVTGSSNRT	NKWNPQNLALS	LHQDRKTNPHDEFNAEEFQ	62	
		: : : : : :	: : : : : :	: : : : : :	: : : : : :	
QY	89	KLDPLAKDVIATLDSQEWPPADK	WGGGLPTAMHISAGTVRIADRGGGTGSQRF	148		
		: : : : : :	: : : : : :	: : : : : :	: : : : : :	
Db	63	KLDYNAIKEDIKLPTLSTQWMP	ADYKCHGGLPTAMHISAGTVRIADRGGGTGSQRF	122		
		: : : : : :	: : : : : :	: : : : : :	: : : : : :	
QY	149	APLNSKPNPDIKDARLLPITKKY	GNKISWADILAGTVAVESMGLPAGVSTGVD	208		
		: : : : : :	: : : : : :	: : : : : :	: : : : : :	
Db	123	APLNSPNDAIKDARLLPITKKY	GNKISWADFLTAGINAVATESMGKTI	182		
		: : : : : :	: : : : : :	: : : : : :	: : : : : :	
QY	209	ITWPEKDIYWGDEKWLAPSDERY	-GDVKKPETMENPLAAYOKGLIYNPVGWNGHPDL	267		
		: : : : : :	: : : : : :	: : : : : :	: : : : : :	
Db	183	WHIPEDYWGDEKWLAPSDERY	SGSD---RELAPDAVOKGLIYNPVGWNGHPDL	236		
		: : : : : :	: : : : : :	: : : : : :	: : : : : :	
QY	268	RTAQGVLETFPAMNDKTAALTAG	HTGVNCHGNSALAPDFKASVQENMGNGEN	327		
		: : : : : :	: : : : : :	: : : : : :	: : : : : :	

Db	183	WHPHEEDVYKSEKWL- ---SERYSGD- ---RELNLPLAAVQNGLYVNPPEGDKPDKP 236
Qy	268	TAAQVLETTFAHMADEKTAALTAGTGHVGNCHNGNAGSALAPDPKASDVENGLDGN 377
Db	237	AAARQRETFRRMGNDSEVALLAGHGFTHGAKHAGPATHVGPPEAPETIAQGLNLS 296
Qy	328	PNMQCKASNAVTSLEGATNPTKFDNGYDYLLEGGYVWELKSPGAAHNEPDIKEN 307
Db	297	SYGKGAGSGTITSLEGATNPTQMDVTSYDMLGTDMLKTSKPAQAQWNAVDPEKD 356
Qy	388	KPYDASDPSIRINPTINDADAKIKNPTFVRAICEKFNADPEYKFKTFAKAWFKLTHRLG 447
Qy	357	LADPADEGSPKPMVMTDOLARDFPEYKIRAFHQNPPEFAFAKAWFKLTHRDG 416
Db	448	PKSRVIGYEPVARDLTHQDIPAGNTDYCE- ---EYFKQIAGSLGSIKSEWSTAMDANT 504
Qy	417	PKTRYLGPVEPKEDFTWQDPIPEVDYELTVEETKIKLINSOLIVSELVYKIMASNT 476
Qy	505	YRSGDMGKGANGARIELAPONEMWGNEPERLAKVLSVYSQSD- ---TCASIAADVLVLAGS 562
Db	477	FRNSCKGKGANGARIELAPOKWENPERLAKVLSYVEDQRELPKSVIADLVLAGS 536
Qy	563	VGLEKAKAGYDVRVYPLAGKGDATKEMTADSFAPLELADGFNNQKKEYVYKPEM 622
Db	537	ANVEAKARAGDVKYEPFPGNDTQEGTDIVESFVLEZFAFPAFNFNQKQZISVYPEEL 596
Qy	623	LIDRAQLMGISYPMVLLAGNARVLTGNYGTHKHGVFTFCBGLTNDGFYNLTDGNSWK 682
Db	597	LVDKAQLGLTAPMTVLLVGLLVGANTRDLPHGVTIDRIGVLTNDEPFLNLDNKEYV 656
Qy	683	PVGSNAYETDRKTAGVMTASRVDLVFGSLSLRSYAEVQAQDDGKEKFRDFAAATK 742
Db	657	PTDSGTDELKQYGEVMTATRVLDLVFGSLSLRSINASTAGETQAQDDQKEKFRDFINAWK 716
Qy	743	VMAHNRDFT 751
Db	717	VMAHNRDFL 725

RESULT 14  
F43892  
F43892 catalase [EC 1.11.1.6] [Imported] - synthetic  
C:Species: synthetic  
C:Name: cat gene engineered and expressed in *Bacillus stearothermophilus*  
C:Date: 24-Mar-2000 #sequence\_revision 24-Mar-2000 #text\_change 23-Mar-2001  
C:Accession: T43882  
C:Author: R.R.Matsuura, T. Miyai, K. Takulnalaenam, S. Yomo, T. Shima, Y. Miki, S. Arita  
C:Title: Evolutionary molecular engineering by random elongation mutagenesis.  
C:Journal: JOURNAL OF MOLECULAR BIOLOGY  
C:Accession: F43889  
C:Volume: 343  
C:Pages: 483-493  
C:Status: translated from GB/EMBL/DBDJ  
C:Molecule type: DNA  
C:Molecule type: RNA  
C:Residues: 1-753 <M>  
C:Cross-references: EMBL:AB020119; NID:g4140145; PIDN:BA037031.1; PID:g4140146  
C:Comment: For the wild-type sequence, see PRJ:JS0520.

[illegible]

QY 149 APLNSHWNDVLDKARLLWPKKKYKNGKLSWADLLIAGVYVYSEMLPAYSSEGRVD 208  
 Db 123 APLNSFWNDKARLLWPKKKYKNGKLSWADLLIAGVYVYSEMLPAYSSEGRVD 182  
 QY 209 INPEKDIYWDSEKWLAPSDRY -GDVNPETMENPLAAVOMGLIIVNPGVNGHDPDL 267  
 Db 183 VHPEDVYVWSGEMKLA--SERVSGD---RELENPLAAVOMGLIIVNPGVNGHDPDL 236  
 QY 268 RTAQOVIETPHAWANDEKTAALTAGCHTVNCHNGNMSALADPPKASDVNGGLWGN 327  
 Db 237 AARD:RETFPRMNGNDEETVALTAGCHTVNCHNGNMSALADPPKASDVNGGLWGN 296  
 QY 328 PRNOKSASNAVSGTEGAMTNPFTKPDGMGVDFLLGVNNELKKSAPAGAHMEPDIKKEN 387  
 Db 237 SYGKGSDTITSGEGAMTNPFTKPDGMGVDFLLGVNNELKKSAPAGAHMEPDIKKEN 356  
 QY 388 KPVKASDSIRHNPWTDADAKIWNVPIATICEKFMADPFFKTKFAKFKIHRDLG 447  
 Db 357 LAPDADSPKAPTNMTTOLALREDFVEKTAARFVGNPEEFAENAFKTHWDG 416  
 QY 448 PKRSYIGVPAEDLLIWDPIAGNTDYCE---EVVKOKIAQGLISISEMVSNTASDART 504  
 Db 417 PKTRYIGVPEKEDDIWDPIDEVDYETAEIEEIKAKILNSGLITSVELKTAASAST 476  
 QY 505 YRGSWNGANGARTALPONONGONERPLAVL SVTEQISAD--TGASIDVIVLAGS 562  
 Db 477 FRNSKGGGANGARIRLAPKDWVNEPRKLVLSYEDIQRELPKVSIADIVLAGS 536  
 QY 563 VGTAKAAAGYDVRVPELKGGRDATAMTDADGSAFPLADGFRNKKQKEYVVKPEEM 622  
 Db 537 AAYEKAAAGDEYVPELKGGRDATAMTDADGSAFPLADGFRNKKQKEYVVKPEEM 596  
 QY 633 LDORALQGLDEPMVULGGMVGLGYNQYGGTKGKGVATDQCOLGNISFVYKNGNSWK 652  
 Db 597 LVDRKALGLAEMVULGGLVGLYANVRLDPGHVFTDRTGLVFNLDWYEMV 686  
 QY 683 PVGSNATYERDRTKAGVKTASRYDVLVFGSNNLSRSTAEVYADONKEKVFADVAATK 742  
 Db 657 PDSGTYEIRDKTEGVMTATRVLDLFGSNNLSRSTAEVYADONKEKVFADVAATK 716  
 QY 743 VKNADRDVL 751  
 Db 717 VKNADRDVL 725

## RESULT 15

F82584

C:Species/perovskite XF2232 [Imported] - Xylella fastidiosa (strain 9a5c)

C:Date: 18-Aug-2000 sequence\_revision 20-Aug-2000 #text\_change 24-Aug-2001

C:Accession: F82584

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: AB2515; PMID:10910347

A:Accession: F82584; Date: 18-Aug-2000; sequence\_revision 20-Aug-2000; #text\_change 24-Aug-2001

A:Status: preliminary

A:Species: preliminary

A:Molecule type: DNA

A:Residues: 1-781 GSIN

A:Cross-references: GB:AE004035; GB:AE003849; MID:99107371; PIDN:AF85031.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A

as-Neto, E.; Docena, C.; El-Dor, J.; Falcão, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froim

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

Chado, M.A.; Madeira, A.M.B.N.; Medeiros, B.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, B.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, N.; Rosa, A.J. de N.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Szwed

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.N.; Silva Jr., W.A.; da Silve

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.N.; Silva Jr., W.A.; da Silve

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.N.; Silva Jr., W.A.; da Silve

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.N.; Silva Jr., W.A.; da Silve

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.N.; Silva Jr., W.A.; da Silve

M.: Teubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
 A:Reference number: A59328  
 C:Accession: A59328  
 C:Genetics: annotation

A:Gene: X2232

C:Superfamily: catalase HPI

C:Keywords: heme; iron; metalloprotein

F:126/Active site: His (distal axial ligand) [status predicted]

F:313/Binding site: heme iron (His) (proximal axial ligand) [status predicted]

F:364/Active site: Trp (cryptophy radical intermediate) [status predicted]

Query Match 56.28; Score 2279; DB 2; Length 781;

Best Local Similarity 56.68; Pred. No. 1.2e-149; Indels 4; Gaps 8;

Matches 431; Conservative 100; Mismatches 176;

QY 40 CPVHGGTST-----GTNKDMPPEGLINLILHQDRKSDPDMPFNYRE 85  
 Db 25 CPTNRPASSETTSKVPFNTAVETGTHKMPNQLRVLDLHQHSNRSPLGETYAK 84  
 QY 86 EYKLDLAKKDLKVALMDSQWPAWDKHYGLMTRMHSACTYATADRGCGGGS 145  
 Db 85 EFQKLDYALAKROLHALMDSQWPAWDKHYGLMTRMHSACTYATADRGCGGAGQ 144  
 QY 146 QRFAPLNSPONSLOKARLLWPKKKYKNGKLSWADLLIAGVYVYSEMLPAYSSE 205  
 Db 145 QRFAPLNSPONSLOKARLLWPKKKYKNGKLSWADLLIAGVYVYSEMLPAYSSE 204  
 QY 206 RVD:WEPEDVYVWSGEMKLA--SERVSGD---RELENPLAAVOMGLIIVNPG 243  
 Db 205 RVD:WEPEDVYVWSGEMKLA--SERVSGD---RELENPLAAVOMGLIIVNPG 243  
 QY 246 AVQNGLIIVNPGVNGHDPDLRTAQOVIETPHAWANDEKTAALTAGCHTVNCH 305  
 Db 264 AVQNGLIIVNPGVNGHDPDLRTAQOVIETPHAWANDEKTAALTAGCHTVNCH 323  
 QY 306 ASALAPKASDVNGGLWGNQYGGTKGKGVATDQCOLGNISFVYKNGNSWK 365  
 Db 324 ADNVGPEPGELEGGGLGWHNRFGSGRAGDITSGLEVTTWTPTQMSDFEHLFYE 383  
 QY 366 WELKSPAGAHMEPDIKKENKPV--DASPSIRHNPIMTDADMAIKVNPYTAICE 423  
 Db 384 WELKSPAGAHMEPDIKKENKPV--DASPSIRHNPIMTDADMAIKVNPYTAICE 439  
 QY 424 MADPEFKTKFAKFKIHRDLGPKSRITGPEVPAEDLIWDPIAGNTDYCE--- 479  
 Db 440 HAHPPQADVFARAFKILHRDNGPVRVYLGPEVPEELIWDPIGVSHVLVDADLL 499  
 QY 480 VKOKIASGLISEMVSNTASRYDVLVFGSNNLSRSTAEVYADONKEKVFADVAAT 539  
 Db 500 LKRSASGGLISQVITMNSSTGTRSGDKGANGRCLAPQSNVQPOLSVLV 559  
 QY 540 SVTEQISADG-----STADIVLAGSVGLERKAAAGYDVRVPELKGGRDATAM 592  
 Db 560 ETLREVTVEFNAQAGDKRISLADIVLAGSVGLERKAAAGYDVRVPELKGGRDATAM 619  
 QY 593 DADSPFAPLADGFRNKKQKEYVVKPEEMLDRAQLMGLTGPEMVLGGHRYGVNG 652  
 Db 620 DADSPFAPLADGFRNKKQKEYVVKPEEMLDRAQLMGLTGPEMVLGGHRYGVNG 679  
 QY 653 GYKRGVTFDCEGGLTNDFVNLDMGNSWPKGS--NAVETDRKTKGAVKATSRVDLV 710  
 Db 680 GYKRGVTFDCEGGLTNDFVNLDMGNSWPKGS--NAVETDRKTKGAVKATSRVDLV 739  
 QY 711 GSNLLSTAEVYADONKEKVFADVAATKVMADRDV 751  
 Db 740 GSNLLSTAEVYADONKEKVFADVAATKVMADRDV 780

Search completed: October 7, 2003, 19:46:43

Job time : 32.6575 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:38:38 ; Search time 75.4005 Seconds  
(without alignments) 2577.085 Million cell updates

**Title:** US-09-884-889-6

perfect score: 4058  
Sequence: 1 MNNASADDLHSSIIQRCRAF.....RDFVAANTKVMNADREDVVAS 753

Scoring table: BLOSUM62

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:

1: sp\_archea:†

2: sp\_bacteria:\*

3: sp\_fungi:★

4: sp\_human:★

5: sp\_invertebrat

6: sp\_mammal:\*

7: sp\_mhc:★

8: sp\_organelle:

9: sp\_phage:★

```
10: sp_plant:*
```

```
11: sp_rodent:*
```

12: sp\_virus:★

13: sp\_vertibrate

14: sp\_unclassified

15: sp\_rv1rug: \*

16: sp bacteriap

17: sp archep: \*

1. *Chlorophyll a* (Chl *a*)

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
1	2920	72.0	721	2	092CM4		Q92M04 Legionella
2	2919	71.9	724	16	Q9KR56		Q9KR56 Vibrio chol
3	2870	70.7	720	2	Q9R6S9		Q9R6S9 synechococ
4	2854	70.3	723	16	Q8D951		Q8D951 Vibrio vuln
5	2820	69.5	731	16	Q8U714		Q8U714 agrobacteri
6	2817	69.4	723	2	Q9R708		Q9R708 agrobacteri
7	2802.5	69.1	728	16	Q8E981		Q8E981 shewanella
8	2795	68.9	754	16	P73911		P73911 synechocys
9	2731	67.3	720	2	055110		055110 synechococ
10	2514	62.0	735	16	Q9KR66		Q9KR66 bacillus ha
11	2489	61.3	743	2	Q9S5K5		Q9S5K5 bacillus st
12	2489	61.3	743	2	Q9S5K0		Q9S5K0 bacillus st
13	2489	61.3	744	2	Q9S5K7		Q9S5K7 bacillus st
14	2489	61.3	744	2	Q9S505		Q9S505 bacillus st
15	2489	61.3	745	2	Q9S504		Q9S504 bacillus st
16	2489	61.3	746	2	Q9S5N4		Q9S5N4 bacillus st

## ALIGNMENTS

[illegible]

Query Match 72.0%; Score 2920; DB 2; Length 721;  
Best Local Similarity 73.3%; Pred. No. 1.8e-192;  
Matches 531; Conservative 77; Mismatches 106; Indels 1

the peroxidases: cloning of the *katB* gene

QY 36 NSGR-----CPVHGNGTSTGSKNDWMPGECINLILHQDRKSDPMDFDNFBREVR 88  
 Db 1 NGRVSGTTCGCVTGGCTATGCTATGAMPALNLDLHNDTRTINPHEADTNREVK 60  
 QY 89 KLDFALAKOHALMTDSCDEWFPADGHWGGLMIRMAHSACTYATADRGCGTGSORF 148  
 Db 61 KLDFEALKOHALMTDSCAMPADGHWGGLMIRMSHAASTYRVADRGAGTGNORF 120  
 QY 149 APLNSHPDNYSDKRLALRIKTKYKNSWADLMLAGTVAYESMGLPAYGFSORVD 208  
 Db 121 APLNSHPDNYSDKRLALRIKTKYKNSWADLMLAGTVAYESMGLPAYGFSORVD 180  
 QY 209 INPEKADLYNGDSEWELAFSDERYGDVKNPPTENPLAAYOMGLIYVNPBGVNGHPDPLR 268  
 Db 181 INPEKADLYNGDSEWELAF--KRY-DKGSRESLENPLAAYOMGLIYVNPBGVNGHPDPLR 237  
 QY 269 TACQVLETFAFAMANDKETAALPACGTYGVNCHGNGNSALAPDPAKASDVENGGLGWNP 328  
 Db 238 TACQVLETFAFAMANDKETAALPACGTYGVNCHGNGNSALAPDPAKASDVENGGLGWNP 296  
 QY 329 NMGKASNAVSTSGEGAMTWPTFKDFMGVFDLLFGYNELKKSAPAGAHMEPIDIKKENK 388  
 Db 298 TTRIGRTNVTSGEGAMTWPTFKDFMGVFDLLFGYNELKKSAPAGAHMEPIDIKKENK 357  
 QY 389 PVDAOSDTHUNPTWDAMATKATVYPAFACKENPADPPTKYTKAFANKLTHRDLP 448  
 Db 358 PVDAOSDTHUNPTWDAMATKATVYPAFACKENPADPPTKYTKAFANKLTHRDLP 417  
 QY 449 KSRVIGPEVPAEDLIWQDPIPAAGNTDYCEVVKQIAQSGLSISEMVSSTAMDSARTYGS 508  
 Db 418 KSRVIGPEVPAEDLIWQDPIPAAGNTDYCEVVKQIAQSGLSISEMVSSTAMDSARTYGS 477  
 QY 509 DMRGANGANIRLAPQWNGNPELAKVLSYQTSISADTCASTADIVLAGSGVETXA 568  
 Db 478 DMRGANGANIRLAPQWNGNPELAKVLSYQTSISADTCASTADIVLAGSGVETXA 537  
 QY 569 AKAAGTVYVPTFAGRGDAETAKTDADSPLEADPFGNMQKKEVYKPEEMILLDRAQ 628  
 Db 538 AKAAGTVYVPTFAGRGDAETAKTDADSPLEADPFGNMQKKEVYKPEEMILLDRAQ 597  
 QY 629 LKGLTGCEMTVLGSRVLGTYGTEKGVPTDCBQZLTNDFFYNTDNGSKWPGVNA 688  
 Db 598 LKGLTGCEMTVLGSRVLGTYGTEKGVPTDCBQZLTNDFFYNTDNGSKWPGVNA 657  
 QY 659 YEIRDRKTAGVKTASRVDLVFGSLSLRSYAETVAQDNGKEKVFDEFAAMTKVNAOR 748  
 Db 628 YEIRDRKTAGVKTASRVDLVFGSLSLRSYAETVAQDNGKEKVFDEFAAMTKVNAOR 717  
 QY 749 FOVA 752  
 Db 718 FDLA 721

## RESULT 2

OSKR55 PRELIMINARY; PRT: 724 AA.  
 ID OSKR55  
 AC 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Catalase/peroxidase.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 RN NCBI\_TaxID=566;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=El Tor N16961 / Serotype O1;  
 RA H30242-20406833; PubMed=10932301;  
 RA H30242-20406833; PubMed=10932301;  
 RA Dodson R.J., Haft D.H., Hickley E.K., Peterson J.D., Umayahara,  
 G.I.I. S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,

RESULT 3

RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,  
 McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,  
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
 cholerae".  
 RL Nature 406:477-483(2000).  
 DR EMBL; AEC04233; AAF94714.1; -.  
 DR HSP; P00431; IBEK.  
 DR TIGR; VCI560.  
 DR InterPro; IPR000763; Bact\_ctase/prase.  
 DR TrEMBL; P00431; PEROXIDASE.  
 DR Pfam; P00458; PEROXIDASE.  
 DR PRINTS; P00458; PEROXIDASE.  
 DR TIGRFAMs; TIGR00198; cat\_per\_HPI; 1.  
 DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
 DR Peroxidase; Complete proteome.  
 SQ SEQUENCE 724 AA; 40650 MW; C53E123D12FD851 CRC64;

Query Match 71.98; Score 2919; DB 16; Length 724;  
 Best Local Similarity 74.18; Prod. No. 2,1e-192;  
 Matches 530; Conservative 75; Mismatches 110; Indels 0; Gaps 0;

QY 37 SGCPVPMGAGTGTSTGSKNDWMPGECINLILHQDRKSDPMDFDNFBREVKIADLAK 96  
 Db 9 SGCPVPMGAGTGTSTGSKNDWMPGECINLILHQDRKSDPMDFDNFBREVKIADLAK 69  
 QY 97 KDVHLMTDSQEWMPADGHWGGLMIRMAHSACTYATADRGCGTGSORFAPLNSWPD 156  
 Db 69 KDVHLMTDSQEWMPADGHWGGLMIRMAHSACTYATADRGCGTGSORFAPLNSWPD 128  
 QY 157 NVSLDKARLLMPTIKKTKYKNSWADLMLAGTVAYESMGLPAYGFSORVDIWEPEKDI 216  
 Db 129 NVSLDKARLLMPTIKKTKYKNSWADLMLAGTVAYESMGLPAYGFSORVDIWEPEKDI 188  
 QY 217 YNGDEKELAPSDERYGDVKNPPTENPLAAYOMGLIYVNPBGVNGHPDPLRTAOVLET 276  
 Db 189 YNGDEKELAPSDERYGDVKNPPTENPLAAYOMGLIYVNPBGVNGHPDPLRTAOVLET 248  
 QY 277 FARMANDEETALTAGGTYGVNCHGNGNSALAPDPAKASDVENGGLGNPMMQKASN 336  
 Db 249 FARMANDEETALTAGGTYGVNCHGNGNSALAPDPAKASDVENGGLGNPMMQKASN 308  
 QY 337 AVTSGLEGATNTPTEDMGVFDLLFGYNELKKSAPAGAHMEPIDIKKENKPVASDPS 396  
 Db 309 AVTSGLEGATNTPTEDMGVFDLLFGYNELKKSAPAGAHMEPIDIKKENKPVASDPS 368  
 QY 397 TEHPHPTDADNAIKVNPYPAFACKENPADPPTKYTKAFANKLTHRDLPKSKRYDPE 456  
 Db 369 TEHPHPTDADNAIKVNPYPAFACKENPADPPTKYTKAFANKLTHRDLPKSKRYDPE 428  
 QY 457 VPAEDLIWQDPIPAAGNTDYCEVVKQIAQSGLSISEMVSSTAMDSARTYGSDMRGANG 516  
 Db 429 VPAEDLIWQDPIPAAGNTDYCEVVKQIAQSGLSISEMVSSTAMDSARTYGSDMRGANG 488  
 QY 517 ARILAPQWNGNPELAKVLSYQTSISADTCASTADIVLAGSGVETXAAMAAAGVDY 576  
 Db 489 ARILAPQWNGNPELAKVLSYQTSISADTCASTADIVLAGSGVETXAAMAAAGVDY 548  
 QY 577 KVPFLKRGDAETAKTDADSPLEADPFGNMQKKEVYKPEEMILLDRAQLMGLTGP 636  
 Db 549 KVPFLKRGDAETAKTDADSPLEADPFGNMQKKEVYKPEEMILLDRAQLMGLTGP 608  
 QY 637 KVPFLKRGDAETAKTDADSPLEADPFGNMQKKEVYKPEEMILLDRAQLMGLTGP 696  
 Db 609 KVPFLKRGDAETAKTDADSPLEADPFGNMQKKEVYKPEEMILLDRAQLMGLTGP 668  
 QY 697 GAVKMTASRVDLVFGSLSLRSYAETVAQDNGKEKVFDEFAAMTKVNAORDFW 751  
 Db 669 GAVKMTASRVDLVFGSLSLRSYAETVAQDNGKEKVFDEFAAMTKVNAORDFW 723

QY	571	AAGDYRVVFLKRGCRDATAEMTDASFAPELPLADGFRNMOKKEYVVKPEMLLDRAQLM	630
DB	539	AAGDYRVVFLKRGCRDATAEMTDASFAPELPLADGFRNMOKKEYVVKPEMLLDRAQLM	598
QY	631	GLTGPEMTVILGGRVRLTNGTGGTKHGVFTDCEGSLTNDNFVNLDMGSKPKVGSNAYE	690
DB	599	GLTGPEMTVILGGRVRLTNGTGGTKHGVFTDCEGSLTNDNFVNLDMGSKPKVGSNAYE	658
QY	691	IDNRKTGAVKVTASRVOLVFVGSNLSILRSYAEVTAODDNGEKFEVDFVAAATKYVNAHDFD	750
DB	659	IDNRKTGAVKVTASRVOLVFVGSNLSILRSYAEVTAODDNGEKFEVDFVAAATKYVNAHDFD	718
QY	751	V 751	
DB	719	L 719	
RESULT 4			
ID	Q80951	PRELIMINARY;	PRT; 723 AA.
AC	Q80951;		
DT	01-MAR-2003 (T-EMBLrel. 23, Created)		
DT	01-MAR-2003 (T-EMBLrel. 23, Last sequence update)		
DT	01-MAR-2003 (T-EMBLrel. 23, Last annotation update)		
DR	Catalase.		
OS	Vibrio vulnificus.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;		
OC	Vibrionaceae; Vibrio.		
OX	NCBI_TaxID=672;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CMCP6;		
RA	Choy H.B.; Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,		
RT	*Complete genome sequence of Vibrio vulnificus CMCP6.;		
RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.		
RM	EMBL: AE016806; AAO1099.1; -.		
SC	Complete proteome.		
SEQ	SEQUENCE 723 AA; 80425 MW; D714DD48FD7073FD CRC64;		
Query Match			
Best Local Similarity 70.3%; Score 2854; DB 16; Length 723;			
Matches 512; Conservative 89; Mismatches 114; Indels 0; Gaps 0;			
QY	37	SGRCPCVHGGTSTGTGSKNDMPGKGLDILHQDQKSDPMDPNTREYRKILDFDA	96
DB	9	SCQPCVHGGATSSNSVAWPAKALNDILHQDQKSNPAGDPTREYRKILDYEAAL	68
QY	97	KOYVAHMTDSQWPAQWGHYGGMLTAMWHSAGTYRATKRGGGTGSGORFAPLNSWD	156
DB	69	RLTKLTMSQWPAQWGHYGGMLTAMWHSAGTYRATKRGGGTGSGORFAPLNSWD	128
QY	157	NVSLDKARLLWPPIKKYGNKISWADLMLAGTVAYESNGLPAYGFSGRVDINPEKDI	216
DB	139	NALDKARLLWPPIKKYGNKISWADLMLAGTVAYESNGLPAYGFSGRVDINPEKDI	188
QY	217	YWGDEKWLAPSDERYGDVKNPENTMENPLAONGLIYNPEGVNGHPDPLRTAQVLET	276
DB	189	YWGDEKWLAPSGAENRSYSGERDLNPAAVTCGLIYNPEGVNGHPDPLRTAKDHVET	248
QY	277	FARMANDEKTAALTAGHTVGNCHGNASALAPDPKASDVNGLNGNPNMKNKASN	336
DB	249	FARMANDEKTAALTAGHTVGNCHGNASALAPDPKASDVNGLNGNPNMKNKSGIRD	308
QY	337	AVTSIGEGAWTTPKTFDMGYFDLLPGYNWELKSPAGAHWPEIDIKENKPVDSAPS	396
DB	309	TVTSIGEGAWTTPMDQNGYFRLMLTNDMLKSPAGAHWPEINPEEDPTVDEPT	368
QY	397	IRHNPIMTDAMAIKNVPTVRAICEKFMADPEYFKTKAKWFKLTHRDIGPKSYIGPE	456
DB	369	IRHNPIMTDAMAIKNVPTVRAICEKFMADPEYFKTKAKWFKLTHRDIGPKSYIGPE	428

QY 457 VPAEDLIWDPPIAGNTDCEVYVVKIAQSGLSISEKVNSTANDSARTYRGSMDRGANG 516  
 Db 429 VPAEDLIWDPPIAGNTDCEVYVVKIAQSGLSISEKVNSTANDSARTYRGSMDRGANG 488  
 QY 517 ARIRLAPONWOGNEPERLAKVLSYDQISADTGASIAADVILAGSVGIEKAAGAAYDV 576  
 Db 489 ARIRLAPONWOGNEPERLAKVLSYDQISADTGASIAADVILAGSVGIEKAAGAAYDV 548  
 QY 577 RYVPLKRGDGTAEKMTDADSFAPLEADGFARMOKKEYVYKPEMLDRAQLMGLTGP 636  
 Db 549 TVPFAAGNDGKTDIDYDVSDFVLEPFAOGRMOKQNTAYNPEMLDRAQLMGLTGP 608  
 QY 637 MYVLGGRHRYLVGTYGKTHGVFTDCBGLTNDFFVNLTDGMSNKPVGSNAEYDRDKT 696  
 Db 609 MYVLGGRHRYLVGTYGKTHGVFTDCBGLTNDFFVNLTDGMSNKPVGSNAEYDRDKT 668  
 QY 697 GAVKNTASRVLDVLPFGSNLSRSLVACVYADONGKEKFRVDPVAAKTYVNNADRFV 751  
 Db 669 GAVKNTASRVLDVLPFGSNLSRSLVACVYADONGKEKFRVDPVAAKTYVNNADRFV 723  
 RESULT 5  
 ID Q80714 PRELIMINARY; PRT: 731 AA.  
 DT 01-JUN-2002 (TREMBLrel, 21, Created)  
 DT 01-JUN-2002 (TREMBLrel, 13, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel, 23, Last annotation update)  
 DE Catalase.  
 GN KATA OR ATU4642 OR AGR\_L\_481.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OX NCBI\_TaxID=176299;  
 RX (1)  
 QY SEQUENCE FROM N.A.  
 RA MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 Chandra V.K., Sheth J., Eisen J.A., Karp P., Bovee D. Sr.,  
 Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri P.,  
 Raymond S., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
 Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 Wester E.W.;  
 RA C58 "genome of the natural genetic engineer Agrobacterium tumefaciens  
 strain C58"  
 RL Science 294:2317-2323(2001).  
 [2]  
 QY SEQUENCE FROM N.A.  
 RA MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 Qurollo B., Gorman B.S., Cho Y., Akenazi M., Halling C., Mullin L.,  
 Nollan C., Allinger M., Doughty J., Sorensen C., Lippert M.,  
 Planegan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
 Cielo C., Slater S.;  
 RA "Genome sequence of the plant pathogen and biotechnology agent  
 Agrobacterium tumefaciens C58."  
 RL Science 294:2323-2328(2001).  
 DR EMBL; AS0008224; AKN68805.1; ALF\_INIT.  
 DR InterPro; IPR000763; Bac.citase/phrase.  
 DR InterPro; IPR002016; Peroxidase.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00141; peroxidase.1.  
 DR PRINTS; PR00458; PEROXIDASE.  
 DR PROSITE; PS00198; cat\_per\_HPI.1.  
 DR PROSITE; PS00478; WD\_REPEATS\_1.  
 DR PROSITE; PS00478; WD\_REPEATS\_1.  
 KW Complete proteome.

SQ SEQUENCE 731 AA; 79570 MW; E1AA67F4E2456F20 CRC64;  
 Query Match: 69.5%; Score 2820; DB 16; Length 731;  
 Residues Identical: 108; Pct. Id.: 14.8%;  
 Matched: 508; Conservative: 86; Mismatches: 121; Indels: 0; Gaps: 0;  
 QY 37 SKCKPVHGGTGTGTSGNKWDVPEGLNLDLHQDQKSDPDPDFTYREVRKLDLQALK 96  
 Db 16 AGKCPVHGGTGTGTSGNKWDVPEGLNLDLHQDQKSDPDPDFTYREVRKLDLQALK 75  
 QY 97 KDYVALTQSDQWPMADNGHYGGLTMAHMSAGTVYIADRGCGGTGSGRFAPLNSWPD 156  
 Db 76 ADLRALMTDSQWPMADNGSYGHWYTHAGSYVTYDGRGNTYGNQRFAPLNSWPD 135  
 QY 157 NVSLDKARLLWPKIKKTKYGNKISNADLILAGTVYVPSMGLPAYGSGRVDIWPBKDI 216  
 Db 136 NVYDKGRHLLWPKIKKTKYGNKISNADLILAGTVYVPSMGLPAYGSGRVDIWPBKDI 195  
 QY 217 YKQDEKHLAPQDEKGVDPKNTPEMLPAAVQNGLIYVWPEVNGHDPDLPRTAQVLET 276  
 Db 196 YKQDEKHLAPQDEKGVDPKNTPEMLPAAVQNGLIYVWPEVNGHDPDLPRTAQVLET 255  
 QY 277 FARMANDEKTAALTAGTGYVCHNGHNSALADPKASDVENOGLNGNPMQCKASN 336  
 Db 256 FARMGDDDEETVALTAGTGYVCHNGHNSALADPKASDVENOGLNGNPMQCKASN 315  
 QY 337 AVYSGTEGATNTYKTFKAYFDLLQYVWELKSPAGAHWFDDIKENKWDADSDPS 396  
 Db 316 TVVSGLEGATSEPTKWDNGFDMLKHEMTLTHSPAGASQAPATIAEDKRPVDEAS 375  
 QY 397 TRHNPIIMTDMAKLVNPTYRAICEKTMADPEYFKTKTAKAMFKLTHRLDLPKSYGTPE 456  
 Db 376 TRTPMWTDMADKLVNPTYRAICEKTMADPEYFKTKTAKAMFKLTHRLDLPKSYGTPE 435  
 QY 457 VPAEDLIWDPPIAGNTDCEVYVVKIAQSGLSISEKVNSTANDSARTYRGSMDRGANG 516  
 Db 436 VPAEDLIWDPPIAGNTDCEVYVVKIAQSGLSISEKVNSTANDSARTYRGSMDRGANG 495  
 QY 517 ARIRLAPONWOGNEPERLAKVLSYDQISADTGASIAADVILAGSVGIEKAAGAAYDV 576  
 Db 496 ARIRLAPONWOGNEPERLAKVLSYDQISADTGASIAADVILAGSVGIEKAAGAAYDV 555  
 QY 577 RYVPLKRGDGTAEKMTDADSFAPLEADGFARMOKKEYVYKPEMLDRAQLMGLTGP 636  
 Db 556 AVPFAAGNDGASQTDADSFAPLEADGFARMOKKEYVYKPEMLDRAQLMGLTGP 615  
 QY 637 MYVLGGRHRYLVGTYGKTHGVFTDCBGLTNDFFVNLTDGMSNKPVGSNAEYDRDKT 696  
 Db 616 MYVLGGRHRYLVGTYGKTHGVFTDCBGLTNDFFVNLTDGMSNKPVGSNAEYDRDKT 675  
 QY 697 GAVKNTASRVLDVLPFGSNLSRSLVACVYADONGKEKFRVDPVAAKTYVNNADRFV 751  
 Db 676 GAVKNTASRVLDVLPFGSNLSRSLVACVYADONGKEKFRVDPVAAKTYVNNADRFV 730  
 RESULT 6  
 ID Q80708 PRELIMINARY; PRT: 723 AA.  
 DT 01-MAY-2000 (TREMBLrel, 13, Created)  
 DT 01-MAY-2000 (TREMBLrel, 13, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel, 23, Last annotation update)  
 DE Peroxidase/catalase (EC 1.11.1.6) [Catalase-peroxidase].  
 GN KATA.  
 OS Agrobacterium tumefaciens.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OX NCBI\_TaxID=358;  
 RX (1)  
 QY SEQUENCE FROM N.A.  
 RA MEDLINE=20118000; PubMed=10652101;  
 RA Fan X.-Q., Fan S.-Q.;  
 RA "An Agrobacterium tumefaciens is a virulence factor involved in

[illegible]











RESULT 15



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 06:25:58 ; Search time 8119.46 Seconds  
11276.082 Million cell updates/sec

Title: US-09-884-889-7

Perfect score: 2238

Sequence: 1 atggaataacacacacactc.....accggtttgatctgaaataa 2238

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

1: gb-ba.\*

2: gb-ba.\*

3: gb-ba.\*

4: gb-ba.\*

5: gb-ov.\*

6: gb-pat.\*

7: gb-ph.\*

8: gb-pl.\*

9: gb-pl.\*

10: gb-pl.\*

11: gb-pl.\*

12: gb-sy.\*

13: gb-un.\*

14: gb-yi.\*

15: en-ba.\*

16: en-ba.\*

17: en-ba.\*

18: en-ba.\*

19: en-ba.\*

20: en-ba.\*

21: en-or.\*

22: en-ov.\*

23: en-pat.\*

24: en-ph.\*

25: en-pl.\*

26: en-ro.\*

27: en-sts.\*

28: en-un.\*

29: en-vi.\*

30: en-htg-hum.\*

31: en-htg-inv.\*

32: en-htg-other.\*

33: en-htg-pld.\*

34: en-htg-rod.\*

35: en-htg-rod.\*

36: en-htg-mam.\*

37: en-htg-vrt.\*

38: en-sy.\*

39: en-htg-hum.\*

40: en-htg-mus.\*

41: en-htg-other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2238	100.0	2238	6	AR098264 Sequence
2	1024.2	47.1	11320	9	AR010752 Sequence
3	1024.2	47.1	11320	9	AR010752 Sequence
c	991.8	44.3	340857	1	AL646061 Ralstonia
c	960	42.9	10578	1	AP003010 Mesorhizo
c	951.2	42.5	311963	1	AE007314 Sinorhizo
c	932.4	41.7	2904	1	AE016872 Pseudomon
8	932	40.8	1657	1	Y14317 Streptomyces
9	932	40.8	1657	1	Y14317 Streptomyces
10	932	40.8	1657	1	Y14317 Streptomyces
11	903.2	40.4	4754	1	AR040244 Burkholder
12	903.2	40.4	4754	1	AR040244 Burkholder
13	903.2	40.4	4754	1	AR040244 Burkholder
c	900.2	40.2	11677	1	SC0939106
c	896.6	40.1	3027	1	AE011760 Xanthom
14	896.6	40.1	3027	1	AE011760 Xanthom
15	890.8	39.8	298900	1	AF126956 Streptomy
c	881	39.4	2899	1	AF486647 Ralstonia
c	875.2	38.9	11476	1	AE012219 Xanthom
c	875.2	38.9	11476	1	AE012219 Xanthom
c	871.6	38.9	11476	1	AE012219 Xanthom
c	865.6	38.7	2232	1	AB020234 Bacillus
21	865.6	38.7	2232	1	AB020234 Bacillus
22	865.6	38.7	2232	1	AB020234 Bacillus
23	865.6	38.7	2235	1	AB020120 Bacillus
24	865.6	38.7	2235	1	AB020067 Bacillus
25	865.6	38.7	2235	1	AB020092 Bacillus
26	865.6	38.7	2235	1	AB020092 Bacillus
27	865.6	38.7	2247	1	AB020078 Bacillus
28	865.6	38.7	2250	1	AB020106 Bacillus
29	865.6	38.7	2253	1	AB020116 Bacillus
30	865.6	38.7	2253	1	AB020117 Bacillus
31	865.6	38.7	2256	1	AB020064 Bacillus
32	865.6	38.7	2256	1	AB020066 Bacillus
33	865.6	38.7	2256	1	AB020072 Bacillus
34	865.6	38.7	2256	1	AB020072 Bacillus
35	865.6	38.7	2256	1	AB020073 Bacillus
36	865.6	38.7	2256	1	AB020074 Bacillus
37	865.6	38.7	2256	1	AB020075 Bacillus
38	865.6	38.7	2256	1	AB020075 Bacillus
39	865.6	38.7	2256	1	AB020077 Bacillus
40	865.6	38.7	2256	1	AB020080 Bacillus
41	865.6	38.7	2256	1	AB020081 Bacillus
42	865.6	38.7	2256	1	AB020082 Bacillus
43	865.6	38.7	2256	1	AB020083 Bacillus
44	865.6	38.7	2256	1	AB020083 Bacillus
45	865.6	38.7	2256	1	AB020084 Bacillus

# ALIGNMENTS

RESULT 1

AR098264

LOCUS

DEFINITION

AR098264

VERSION

AR098264.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 2238)

AUTHORS

Robertson,D.E., Sanyal,I. and Adhikary,R.S.

TITLE

Catalases

FEATURES

AR098264 7 from patent US 6074860. DNA linear PAT 14-FEB-2001

Sequence AR098264

AR098264.1 GI:12807521

Unknown.

Unclassified.

1 (bases 1 to 2238)

Robertson,D.E., Sanyal,I. and Adhikary,R.S.

Catalases

Patent: US 6074860-A 7 13-JUN-2000;

Location/Qualifiers

source	1. 2238	634 a	545 c	605 g	454 t
BASE COUNT					
ORIGIN					
Query Match	100.0%;	Score 2238;	DB 6;	Length 2238;	
Best Local Similarity	100.0%;	Pred. NO. 0;			
Matches 2238;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
1	ATGGGAATTCAGCAACCTCAGGATCTCTACGATATACCAAACTCGGCGGAATGC	60			
Db	1 ATGGGAATTCAGCAACCTCAGGATCTCTACGATATACCAAACTCGGCGGAATGC	60			
Qy	61 CTTTTCACCGAGGTTCGGCTTAAGCAAGTGCAGGTGGCGGCACCAAAACAGGATGG	120			
Db	61 CTTTTCACCGAGGTTCGGCTTAAGCAAGTGCAGGTGGCGGCACCAAAACAGGATGG	120			
Qy	121 TGGCCCAATGCTCAGCTCAGCTGCTGATGCGGACCAATTCATGGTATCGGACCAAC	180			
Db	121 TGGCCCAATGCTCAGCTCAGCTGCTGATGCGGACCAATTCATGGTATCGGACCAAC	180			
Qy	181 GACCCGGATTGTACTATCGCGAAGAGTTTAAAGACTAGATCTGGCAGCGTTAAAGAG	240			
Db	181 GACCCGGATTGTACTATCGCGAAGAGTTTAAAGACTAGATCTGGCAGCGTTAAAGAG	240			
Qy	241 GACCTGGCAGCTGATGACAGAGTACAGAGCTGCTGGCAGCAGATAGCTGATAT	300			
Db	241 GACCTGGCAGCTGATGACAGAGTACAGAGCTGCTGGCAGCAGATAGCTGATAT	300			
Qy	301 GGGCCCTCTTTATACGATGGGCTGGCAGCGCGGCACCTACGCTATCGGTGATGGC	360			
Db	301 GGGCCCTCTTTATACGATGGGCTGGCAGCGCGGCACCTACGCTATCGGTGATGGC	360			
Qy	361 CQTGGTGGCGGTGGCTCGCGCTCCACAGCGCTTCGCGCCTCTCAATAGCTGCCAGACAT	420			
Db	361 CQTGGTGGCGGTGGCTCGCGCTCCACAGCGCTTCGCGCCTCTCAATAGCTGCCAGACAT	420			
Qy	421 GCGAATCTGGATTAAGAGAGCTTGGCTTCTTTGGCGCATCAAAACAAAATAGCTCGAAA	480			
Db	421 GCGAATCTGGATTAAGAGAGCTTGGCTTCTTTGGCGCATCAAAACAAAATAGCTCGAAA	480			
Qy	481 ATCTCTGGCGGATATGATGATCTACAGAGAACAGTACTCTGGAACATATGGCGTTT	540			
Db	481 ATCTCTGGCGGATATGATGATCTACAGAGAACAGTACTCTGGAACATATGGCGTTT	540			
Qy	541 AAACCTTGGCTTTCAGTGGCGAGACAGATGATGGAACCTCGAAGAGATGATAC	600			
Db	541 AAACCTTTCGTTTGGAGTGGAGACAGATGATGGAACCTCGAAGAGATGATAC	600			
Qy	601 TGGGAGCAGAAACCGATGGCTGGGAGCAAGCGCTATGAAGTGACGAGAGCTCGAA	660			
Db	601 TGGGAGCAGAAACCGATGGCTGGGAGCAAGCGCTATGAAGTGACGAGAGCTCGAA	660			
Qy	661 ATGCGCTGGGAGCGCTCAATATGGGACTCATCTATGTAACCCGAGAGCCACAGCGC	720			
Db	661 ATGCGCTGGGAGCGCTCAATATGGGACTCATCTATGTAACCCGAGAGCCACAGCGC	720			
Qy	721 AAGCCAGACCTATCGCTCTCGCGGTGATATGCTGAGAGACTTTTGGCCGAATGGCAAGT	780			
Db	721 AAGCCAGACCTATCGCTCTCGCGGTGATATGCTGAGAGACTTTTGGCCGAATGGCAAGT	780			
Qy	781 ATGACAGCAAAACCGTGGCTCTCATATGAGCGGTGGACACCTTCGGAAAACCCATGGT	840			
Db	781 ATGACAGCAAAACCGTGGCTCTCATATGAGCGGTGGACACCTTCGGAAAACCCATGGT	840			
Qy	841 GCTGCCATTCGGAGAAATATGTGGCGGACAGAGCTCGCCGCGAGCTATTGAGAAATG	900			
Db	841 GCTGCCATTCGGAGAAATATGTGGCGGACAGAGCTCGCCGCGAGCTATTGAGAAATG	900			
Qy	901 AGCGTGGGTGGAAACAGCTGACGACCCGAGTCCGGATACCATCGACATCGATGGA	960			
Db	901 AGCGTGGGTGGAAACAGCTGACGACCCGAGTCCGGATACCATCGACATCGATGGA	960			

Qy	961 CTAGAAGCGCCTGGCAGCAAGCCCTACTCAATGGAGCAATAACTTTTTTGAACCTC	1020
Db	961 CTAGAAGCGCCTGGCAGCAAGCCCTACTCAATGGAGCAATAACTTTTTTGAACCTC	1020
Qy	1021 TTGGTTCAGAGCTGGAGCTTACCAAAAGTCCAGCTGAGCTTATGAGTGAACCAAA	1080
Db	1021 TTGGTTCAGAGCTGGAGCTTACCAAAAGTCCAGCTGAGCTTATGAGTGAACCAAA	1080
Qy	1081 GACGGTGGCGGGCTGGCAGCAATCCGATGACATGATCCAGCAAGTCGCAAGCTCA	1140
Db	1081 GACGGTGGCGGGCTGGCAGCAATCCGATGACATGATCCAGCAAGTCGCAAGCTCA	1140
Qy	1141 TTATGCTCACTAGTGGAGCTCGCGCTGCGATGAGCTGATGAGCAAAAATTTTCGA	1200
Db	1141 TTATGCTCACTAGTGGAGCTCGCGCTGCGATGAGCTGATGAGCAAAAATTTTCGA	1200
Qy	1201 CGGTACTATGAAACCCCTGATGAGTTTGCAGATGCTTCGCGCAAGCAATGGTACAACTG	1260
Db	1201 CGGTACTATGAAACCCCTGATGAGTTTGCAGATGCTTCGCGCAAGCAATGGTACAACTG	1260
Qy	1261 ACACACAGATATGGCCACCAAGTGGCTACCTGGCGACAGAGTGGCTAGGACAG	1320
Db	1261 ACACACAGATATGGCCACCAAGTGGCTACCTGGCGACAGAGTGGCTAGGACAG	1320
Qy	1321 CTCATCTGGCAGACCTTATACCAAGTATAGCCATCTCTGTAGAGCAAAAACGATAT	1380
Db	1321 CTCATCTGGCAGACCTTATACCAAGTATAGCCATCTCTGTAGAGCAAAAACGATAT	1380
Qy	1381 GAAGGCTTAAAGCCAAATCTCGAATCGGCACTGACGCTAGCGAGCTGTAGGACG	1440
Db	1381 GAAGGCTTAAAGCCAAATCTCGAATCGGCACTGACGCTAGCGAGCTGTAGGACG	1440
Qy	1441 GATATGGCTTCGCATCTACTTTTAAAGACTTCGACAGCGCGGCTGCCAACGGTGA	1500
Db	1441 GATATGGCTTCGCATCTACTTTTAAAGACTTCGACAGCGCGGCTGCCAACGGTGA	1500
Qy	1501 CQTATACGACTGGCCCCACAAAGACTGGGAGCTAAACAACTCTAGCAACTTGGCAG	1560
Db	1501 CQTATACGACTGGCCCCACAAAGACTGGGAGCTAAACAACTCTAGCAACTTGGCAG	1560
Qy	1561 GTACTCANNACACTAGAGGATATCCAGAGGACTTTTACCGGCGCATTCAGATACAA	1620
Db	1561 GTACTCANNACACTAGAGGATATCCAGAGGACTTTTACCGGCGCATTCAGATACAA	1620
Qy	1621 GCAGTATGCTTGGCGACCTGATTTGCTGGCGGCTCGCGGTGTAGAAAAGCTGCA	1680
Db	1621 GCAGTATGCTTGGCGACCTGATTTGCTGGCGGCTCGCGGTGTAGAAAAGCTGCA	1680
Qy	1681 AAGATGCTGGCAGCTGCTGAGAGCTGCTTACCGGCGAGCGGATGCCACGCT	1740
Db	1681 AAGATGCTGGCAGCTGCTGAGAGCTGCTTACCGGCGAGCGGATGCCACGCT	1740
Qy	1741 GAGCAACCGATGTGGAGCTTTCGAAGCACTAGAGCAGCGCGCTGACGCTTTAGAAAC	1800
Db	1741 GAGCAACCGATGTGGAGCTTTCGAAGCACTAGAGCAGCGCGCTGACGCTTTAGAAAC	1800
Qy	1801 TACATTTAAACCGAGCATTAAGATTCGCTGAGCAAAATGCTGTGAGACCGGCGACCT	1860
Db	1801 TACATTTAAACCGAGCATTAAGATTCGCTGAGCAAAATGCTGTGAGACCGGCGACCT	1860
Qy	1861 CTGTGCTTTTCGGCACCAAGAAATGACTGCTTGGTAGGCGGTATGGGTACTGGGCAAC	1920
Db	1861 CTGTGCTTTTCGGCACCAAGAAATGACTGCTTGGTAGGCGGTATGGGTACTGGGCAAC	1920
Qy	1921 AACTACGACGCTGCGACATGCGCTTTACAAATATGCGCTGACCTATCGATGAC	1980
Db	1921 AACTACGACGCTGCGACATGCGCTTTACAAATATGCGCTGACCTATCGATGAC	1980
Qy	1981 TTCTTTGTAAACCTGCTAGACTTCAACACTTAATGGCGACGACGATGAATCAGACAA	2040
Db	1981 TTCTTTGTAAACCTGCTAGACTTCAACACTTAATGGCGACGACGATGAATCAGACAA	2040
Qy	2041 GTTTTGTGAGCGAGAGACTTCAAACTGGCGAGTAAAGTGGAGTGGCAGCGGCTAGAC	2100

```

Db 2041 GTTTTGAAGCGAGAGACTTCAAACTGGCGAAGTAAGTGGAGTGGCCAGCCGGGTAGAC 2100
Oy 2101 CTGATCTTCGATCCCAATCCGAGCTTAAGAGCCCTCGAGAGTCTACGGCTGTGCAGAT 2160
Db 2101 CTGATCTTCGATCCCAATTCGAGAGCTTAAGAGCCCTCGCAGAGTCTACGGCTGTGCAGAT 2160
Oy 2161 TCTGAGAAAGATTGTGTAAGATTTTCTGAAGCCCTGGGCCAAGGTAATGACCTGGAC 2220
Db 2161 TCTGAGAAAGATTGTGTAAGATTTTCTGAAGCCCTGGGCCAAGGTAATGACCTGGAC 2220
Oy 2221 CGGTTTGATCTGAATAA 2238
Db 2221 CGGTTTGATCTGAATAA 2238

RESULT 2
LOCUS AR215252 2238 bp mRNA linear PAT 25-SEP-2002
DEFINITION Sequence 7 from patent US 6410290.
ACCESSION AR215252
VERSION AR215252.1 GI:23313419
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2238)
AUTHORS Robertson,D.E., Sanyal,I. and Adhikary,R.S.
TITLE Catalases
JOURNAL Patent: US 6410290-A 7 25-JUN-2002;
FEATURES
location/Qualifiers
source 1..2238
BASE COUNT 634 a 345 c 805 g 454 t
ORIGIN

Query Match 100.0%; Score 2238; DB 6; Length 2238;
Beat Local Similarity 100.0%; Pred. No. 0;
Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ATGGAAATTCACAAACACTCAGAGTCTCTACGTATACACAAACACTGGCGAAATGC 60
Db 1 ATGGAAATTCACAAACACTCAGAGTCTCTACGTATACACAAACACTGGCGAAATGC 60
Oy 61 CTTTATACGGAGTTCGTTTAAAGATTCAGCAAGTGCAGGTGGCGCCACCAAAACAGGATTCG 120
Db 61 CTTTATACGGAGTTCGTTTAAAGATTCAGCAAGTGCAGGTGGCGCCACCAAAACAGGATTCG 120
Oy 121 TGGCCGACATGCTCAACATGCTTACGCAATTCACGCAATTCATCTGCTGCGAGCCCAAC 180
Db 121 TGGCCGACATGCTCAACATGCTTACGCAATTCACGCAATTCATCTGCTGCGAGCCCAAC 180
Oy 131 TGGCCGACATGCTCAACATGCTTACGCAATTCACGCAATTCATCTGCTGCGAGCCCAAC 180
Db 131 TGGCCGACATGCTCAACATGCTTACGCAATTCACGCAATTCATCTGCTGCGAGCCCAAC 180
Oy 181 GACCCGATTTTACCTATGCGAAGATTTAAGAACTAGATCTGGCAGCGGTTAAAG 240
Db 181 GACCCGATTTTACCTATGCGAAGATTTAAGAACTAGATCTGGCAGCGGTTAAAG 240
Oy 241 GACCTGCGAGCGCTAATGACAGATTCACAGAGCTGTGGCGACGAGATTACGGTCTAT 300
Db 241 GACCTGCGAGCGCTAATGACAGATTCACAGAGCTGTGGCGACGAGATTACGGTCTAT 300
Oy 301 GCGCCCTCTTTATACGATGCGCTGGCAGCGCGCGCCTACCGTATCGSTGATGGC 360
Db 301 GCGCCCTCTTTATACGATGCGCTGGCAGCGCGCGCCTACCGTATCGSTGATGGC 360
Oy 361 GCGCGCTCTTTATACGATGCGCTGGCAGCGCGCGCCTACCGTATCGSTGATGGC 360
Db 361 GCGCGCTCTTTATACGATGCGCTGGCAGCGCGCGCCTACCGTATCGSTGATGGC 360
Oy 421 GCAATCTCGAATAAGCAGCTTCTCTTTTGGCCCATCAACAAATAACGGTTCGAGAA 480
Db 421 GCAATCTCGAATAAGCAGCTTCTCTTTTGGCCCATCAACAAATAACGGTTCGAGAA 480
Oy 481 ATCTCTGGCGGAGTCTATGATGATCTACAGGAACGTAGCTCTGGAACTATGGGCTTT 540

```





/gene="MA0970"	Db	9070	GGGAGGACTTCAATTACCGAAGAGGTTCAAGAGCTAGATCTGGCGCTGTGAAAAAG	1129
/codon_table=11	Qy	242	ACTGTGGAGCGGCTAATGACAGATTCACAGGACTGGTGGCAGCAGATTAAGCTATTATG	301
/product="sensory transduction histidine kinase"	Db	9130	ATCTTGGCGGCTGATGACCGATTCGAGBACTGGTGGCGCGGATTTGGCCATTATG	9189
/db_xref="GI:19914795"	Qy	302	GGCCCTCTTTTATACGATGGGTGGCAGACGGCGGCACTTACCTAGCTGATGGGTGATGGCC	361
/translacion="MOEYKINSMSKQIKGNNAKVEPTATNOLVLSVDNNGTI	Db	9190	GGCTTTTATTTTATTCGATGGCATGGCAGACAGGAGCTATATTCGCCGGTGGTGAAGCCGC	9249
INSNASEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	362	GTGGTGGCGGCTCGCGCTCANGCGTTTCGCGCTCTCAATCTTGGTGGCCAGACATG	421
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9250	GGCGGGCGGCGGTAGAGCGCAGACGCTTTCGCGCGCTCAACAGTTGGCGGACAAATG	9309
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	422	CCAAATCTGGATTAAGACAGCTTGCTTTTGGCCCATCAACAAATAATAGCGTTCGAAAA	481
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9310	TCACCTTGGCAAGACGGCTGCTCTGTGGCGCATGACGACAGATATGGCCGGAAAA	9369
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	482	TCCATCTGGCAAGACGGCTTAAATGATCTACTGACGAAGAGCTCTGGAACATGGCGCTTA	541
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9370	TTCATGGCGGACCTTATGATCTTACGGGCACAGCTGCATGCGAATGAGAACTATGSGATTCA	9429
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	542	AAACTTTTGGTGGATGAGAGAGATGATGGAGCGCTGAAGAAGATGATATCT	601
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9430	AAAGCTTTTTCGGGGCGGTTCGCGAGAGCTCTGGAGCGCCGACAGAGCTGATTT	9489
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	602	GGGAGCAGAAACCGAATGATCTACTGATGAGAGTGGAGCTTGAAGTGGACGAGAGCTCGAAA	661
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9490	GGGGCTCTGAGACACGCTGGCTGATGAGAGCTACACGGCGACCGGAGCTCGAAA	9549
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	662	ATCCCTCTGGAGCGGCTCAAAATGGAGTCACTCTGTAAACCCGAGGAGACCAACAGCA	721
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9550	TCTCGCTCGCTCGCTGAGTGGGGTGTATCTATGATCTCATCCGGAAGGCGCGAACCGCA	9609
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	722	AGCCGATCTATCGCTGCTGGCGCTGATCTGAGAGCTTTTGGCGAATGGCGATGA	781
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9610	ATCCGGATCCGATGAGCGCGGCCAAAGATATCGCGAGGTTTTCGTGCGATGCCCATGA	9669
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	782	ATGAGAAAGAACCGTGCTCTATACGGGGTGGACAGCTTCGAAAAAACCATGGTG	841
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9670	AGCAGAGAGAGCTTGGCTCTTCGCTGGGGTTCAGCTCTCGAAGAACCCAGGTG	9729
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	842	CTGCGATGCGGAGAAATATGTGGCGGAGAGCTGCGCGCAGAGTATTGAAGAATGA	901
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9730	CGGCGCGCTCG--GTCCATGCTCGGGCTGAGCGCAGAGCGCCAGTATGAGCGCGCAAG	9786
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	902	GGCTTGGGGTGGAAAAACCTACGCGCACCGACAGCGTGGCGGATACCATCACTAGTGAC	961
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9782	TATCGCTTGGAGAGCTTCCGCTATGCGAAGCGATGACACCTACCGCGGCGC	9846
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	962	TAGAGCGGCTGGACCAAGACCTACTACTAGGAGCATGATTTTTTGAAGAACTCT	1021
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9847	TGAGAGTCACTTGGACCAACCAACCAAGAGTGAAGTAACTCTTCAGGATCTGT	9906
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	1022	TGCTTGAAGATGGAGCTTACAAAGTCCAGTGGAGCTTATCAGTGGAAACCAAG	1081
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9907	TCGGCTACAGATGGGAATGACAGAGCCCGCGCTGCATCTCAGTGGAAACCTAAG	9966
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	1082	ACGTTGGCGGGCTGGCCACATACCGGATGCACATGATCCAGCAAGTGCAGCGCTCAT	1141
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	9967	GTGGCGCAGTGGCGGACAGTTCGGGAGCGCCACGACCTTCCAGGCTACGCGGCCAA	10026
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	1142	TATGCTCACTACGAGCTTGGCGCTGGCGATGAGCCCTGATATGAGAAAAATTTCTGAC	1201
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	1202	GATATGAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1261
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	10087	ACTTCTAGAGAACCGGGGTCAATATGAGCGGACTCATTCGCCGGCGATGGTCAAGTGA	10146
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Db	1262	CACACAGATATGAGCAAGAGTGGGTACTCTGGAGACAGAGTGGCTCAGAGACAG	1321
HTSDEPILKMGVSGVSKGLPSSIVOLVORVSRNSPEKTEVKNGLYLIVVFSPL	Qy	10147	CTACCGCGACATGGTTCGGCGCGCCCTTCTCTGGGTTCGGAGTTCCTGGAGAGAC	10206

BASE COUNT 3077 a 2495 c 2699 g 3049 t

ORIGIN

Query Match 47.1%; Score 1053.2; Db 1; Length 11320;

Best Local Similarity 68.2%; Pred. No. 2.6e-273;

Matches 1479; Conservative 0; Mismatches 688; Indels 3; Gaps 1;

Qy 62 CTTTACCGAGGTCGCTTAAAGAAAGTGCNGTGGCGGCAACCAAGAGCTGCT 121

Db 8950 CTGTAAAGCGGAGCAACCAACAGAAACCGGTAGAGACATGTCGAAACGCGACTGCT 9009

Qy 122 GGGCCCAATCTCACTCGGATCTTACGCCCAACATTCATCGCTATCGGACCCCAAGC 181

Db 9010 GGGCCCAATCTTGAAGCTTGAAGTCTTCCACCGAGATCTTCCAGTGCACCGGATGG 9069

Qy 182 ACCCGGATTTTCGACCGAGAGTGTACAGCTAGCTAGCTTGGAGTGGCTCAGAGACAG 241

QY	1322	TACTGCGCAGACCCCTATACACAGATGTAGACGATCCCTGTGTAGACGAACGATATTC	1381
DB	10207	TGACTGCGCAGACCCCTATACACAGATGTAGACGATCCCTGTGTAGACGAACGATATTC	10266
QY	1382	AGGCGCTAAAGCGCAAACTCTGGTAATCGGCACTGAGCGTAAAGCGAGCTGGTAAAGCAGG	1441
DB	10267	CCTTCCTCAAGAGAGATCTCTGCGTCAGGCGCTGTGCATCTCGCAACTGGTTCGACCG	10326
QY	1442	CATGGCGCTGTGCATCTCTTATAGAACTGTGTAGAGCGCGCGGTGCGACGCGCAC	1501
DB	10327	CTGTCGCGCTGTGCATCTCTTATAGAACTGTGTAGAGCGCGCGGTGCGACGCGCAC	10386
QY	1501	GTATACGACTGGCCCGCAAAAAGATGGGAGTAACCACTCAGCACTCTGCCACG	1561
DB	10387	GAATTCGCTGGCGCGCAGAGAGATGGGAGTAACCACTCAGCACTCTGCCACG	10446
QY	1562	TACTCAAACTATAGATGTATCGAGAGAGATTAACAGCGCGCATCATGATACAAAG	1621
DB	10447	TGCTGAACACTCTGAGGCTATCCAGACGAGTTCACAGCGCGCGCTCCGCTGCGCAGA	10506
QY	1622	CAGTATCGTGGCGGCACTGATTGTGCTGGCGCGCGCTGCGGGTGTAGAAAAGCTGCA	1681
DB	10507	AGATATCGTGGCGGCACTGATTGTGCTGGCGCGCGCTGCGGGTGTAGAAAAGCTGCA	10566
QY	1682	AGATATCGTGGCGGCACTGATTGTGCTGGCGCGCGCTGCGGGTGTAGAAAAGCTGCA	1741
DB	10567	AGATATCGTGGCGGCACTGATTGTGCTGGCGCGCGCTGCGGGTGTAGAAAAGCTGCA	10626
QY	1742	AGCAACCGATGTGGAGCTTTTCCAGCACTAGACCGCGCTGACGCGCTTTAGAAACT	1801
DB	10627	ACGACCGATGTGGAGCTTTTCCAGCACTAGACCGCGCTGACGCGCTTTAGAAACT	10686
QY	1802	ACATTAACCGGAGATAAGTATCCGCTGAGGAAATGCTGTACAGCGCGCGCTGCTC	1861
DB	10687	ATCTCAAGCGCCAGTACCCCTTACCGCGCGCGAGTTCGTGTTGACACAGCGCAATTGC	10746
QY	1862	TGTCGCTTTCCGCGCACAGAAATGACTGCTTTGGTAGCGCGTATCGGTGACTTGGCACA	1921
DB	10747	TGACTCTGACCGACAGAGATGACGCTGCTTTGGTAGCGCGTATCGGTGACTTGGCACA	10806
QY	1922	ACTACGACGCTTCCGCGCAGTGGCTGTTACAAATAGCGCGGTGACCTATCCATGACT	1981
DB	10807	ACTTGGCAGATACCCAGCGCGCTTTTACCCAGAGCGCGAAGCGCTTAACCAAGACT	10866
QY	1982	TCTTTGTAACCTGCTAGACCTCAACACTAAATGGGAGCCAGCGCATGAATCAGACAAG	2041
DB	10867	TCTTTGTAACCTGCTAGACCTCAACACTAAATGGGAGCCAGCGCATGAATCAGACAAG	10926
QY	2042	TTTTTGAAGCGCAGAGCTCAAACTGCGGCAAGTAAAGTGGAGTGCCACCCGGTGACC	2101
DB	10927	TATTTGAAGGCTCGCATGCGCAAGCTGCGCAAGTCAAGTGGACCGCGCGGTGTCATC	10986
QY	2102	TGATCTTGGGATCCATTCGAGCTTACAGCGCTCGGAGAGTGTACGCGTGTGCGAGT	2161
DB	10987	TGATCTTGGGATCCATTCGAGCTTACAGCGCTCGGAGAGTGTACGCGTGTGCGAGT	11046
QY	2162	CTGAAGAAGATTTGTTAAAGATTTTGTGAAGCTTGGCGCCCAAGTATGAGCACTGACC	2221
DB	11047	CCGAGAGAAATTTGTTCAAGCACTTGTGGCGCGCTGCGCCCAAGTATGAGCACTGACC	11106
QY	2222	GCTTTGACT 2231	
DB	11107	GCTTCGACT 11116	
RESULT 4			
AL646061/c			
LOCUS			
DEFINITION			
Rai1stonia solanacearum GN11000 chromosome, complete sequence;			
ACCESSION			
AL646061			
VERSION			
AL646061.1 GI:17427781			

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE







```

LAISPAWMEVTFWFMVGPGATIGVPLIVVYNCAHEQSTRIMATILSGRQST
10712..11839
/gene="ml16729"
CDS
10712..11839
/gene="ml16729"
/codon_start=1
/transl_table=11
/protein_id="BAB52965.1"
/protein_len="11839"
/db_xref="GI:14026367"
/translation="MANAPAGMNAHALEPPANPAEPLTVPEOTIPICPESTGCD
LPKQDPAMRQSFAMLMDFVEPTGAEFGAGTQIMDLGLDALAUNGSLDFTSLA
LPRQDPDMIVLTLDQGRSTQITPFSKFGDAGSVQHPHGFVEFVSITSEMILL
FVPTDCDSMGLAAAFBSKVGCMGCPILLADLASLVQRPLPMGYNOLPGLISATRA
LGLLVASGVPLFVFGVRRMSDVGNTMVLSTGTFGLAGLVLTGWPESTGLWVGLIGL
RLEFASGVTVHLCQHRNLTGGLVPLFVFGVRRMSDVGNTMVLSTGTFGLAGLVLT
complement(12495..15662)
/gene="ml16730"
complement(12495..15662)
/gene="ml16730"
/codon_start=1

Query Watch
Best Local Similarity 44.3%; Score 991.8; DB 1; Length 340857;
Matches 1429; Conservative 66.8%; Protein 26.8%;
Mismatches 707; Indels 3; Gaps 1;

QY 90 TGCAGGTGGGGCCACCAAAAGGAGTTGGTGGCCCAACATGCTCAACCTCGGCATCTT 149
Db 17392 TGGATCCCGGTGGAGCAACCCGATATTTGGTGGCCCAACGATCGACATCTT 17451
QY 150 AGCCCAACNTATCAGCTATCGGACCAACCAAGCAGCCGATTTGACTATGCGAGAGTT 209
Db 197452 GCATCAACCATCTCTCTGCTGCCACCGCATGGTGAGGCGGCTTCGACTATGCCGGAATT 17511
QY 210 TAGAAGCTAGATCTGGCAGCGGTTAAAAGGAGCTGGCAGGCTAATGACAGATTCCAA 269
Db 197512 CAGAGCTCGACCTACCGCATCTCAAGGAGACCTGGCAGCGGCTGATGACGATTCGCA 197571
QY 270 GGAATGGGGCCAGATACAGCTCATATGCGCCCTCTTATPACGATACGATTCGCA 329
Db 197572 GGAATGGGGCCAGATATTCGCGCACTACGCGCCGCTCTTCATCGGATGGCCTGGCA 197631
QY 330 CAGCCCGGACCTACCGCTATCGGTGATCGGCGGCTGGTGGCGGCTCCGCGCTCAAGAG 389
Db 197632 CAGCGGAGCAGCTACCGCATCTTCGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 197691
QY 390 CTTGCGGCGCTCTCAATACGTGGCCGACGATGCGCAATCTGATTAAGCAGCGCTGTCT 449
Db 197692 TTTGCGGCGCTCTCAACGATGGCCGACATGTCACGCTGACAGGCGCGCGCGCTTTT 197751
QY 450 TTGGCCCATCAACAAAATACGSGTCGAAAATCTCTTGGCCGGATCTAATGATATCAAC 509
Db 197752 GTGGCGCATCAGCAAGATATGGCCGCAAGATCTCTTGGCGCGACCTCTGATGCTCAC 197811
QY 510 AGGAAGCTAGCTCGAAATCTCGCTTTTAAAGCTGCTTTAAAGCTGCTTTCAGCGTGGAG 569
Db 197812 CGGCAAGCTCGGCTGGAATCGATGGCTCTCAGACCTTTTGCTTTTGGCGGCGCGCG 197871

```

```

QY 570 AGATGTATGGGAGCCTGAAGAAGATGTATATCTGGGAGCAGAAACCGAATGGCTGGGAGA 629
Db 178772 CGACGCTGGGAGCCTGACGAGACGCTTPTACTGGGTCCTCGAAGGCAAGTGGTGGCCGA 197931
QY 630 CAAGCGCTATGAGCTGACGAGAGCTCGGAANTCCCTCGGAGCGAGCTGACAAATGGCACT 689
Db 197932 CGAGCGCTACAGCGGCGAGCCGTGACTGTCAGAACCCGCTGGGCGCGTGCAGATGGGCT 197991
QY 690 CATCTATGTAAACCCGAGAGACCAACGCGAAGCCAGACCTTACTGCTGCTGGCGGTGA 749
Db 197992 GATCTACGTCATCCGAGGAGGCGMCGCAATCCGGATCCGCTGGCGCGGCGCGCGCA 198051
QY 750 TATTCGTGAGACTTTTGGCGGCAATGSCAATGATGACGAGAAACGCTGCTCTCATPAG 809
Db 198052 TATCAGGAGACACATTCGCGCGTATGGGCAATGAAACGACGAGAAACGCTAGCACTCATCG 198111
QY 810 GCGTGGGACACACTTTCGGAAAAACCATGCTGCTGCCGATGGGAGAAATATCTGGGCGG 869
Db 198112 CGGCGGCTACGCTTCGCGAGACCATGCTGCGGCTGAGCG---GAGCTGGTGGGTGT 198168
QY 870 AGAGCTCGCGCGAGGTATGAGAAATGAGCTGGGTGGAAACACTCTAGCGAC 929
Db 198169 CGAGCCGGAAGGCGCTGATATCGGACGACGAGGCGCTTGGCTGGCGAGCAATTCGGCAC 198228
QY 930 CGGACAGGTGGGATACCATCACAGTGGACTAGAGAGCGGCTGGACCAAGACCCCTAC 989
Db 198229 CGGCAAGCGGTTGACCCATCGCGCGGCTTGGAGTCAATTTGGACGACGACCGAC 198288
QY 950 TCAATGGAGCAATAACTTTTTTGAACACTCTTTGGTTACGAGTGGGAGCTTACCAAAAG 1049
Db 198289 CAAATGGAGCAACAACCTTCTTGACACATCTGTTGCGCTTCGACTGGGAACCTGACCAAGAG 198348
QY 1050 TCACGCTGAGCTATGACTGGAAACCAAAAGAGGTGCGCGGCTGGGACCAATACCGCA 1109
Db 198349 CCGCGCAGCGCGCATATGACGCGAGCGAGCGGCTGCGCGGCTGCGGACGCTGGCA 198408
QY 1110 TGCAATATGCCAGCAATGGACGCTCCATTTATGCTCACTACGAGCTCGGCGCTGG 1169
Db 198409 CGCAACACTGGCCAGCGCTCAGGACCCCTCCATGCTGACGACGACCTCGCGCTGG 198468
QY 1170 CATGCGAGCTGATAGCAAAAATTTCTCGAGGTTACTATGAAAACCTGATGAGTTTGC 1229
Db 198469 TTTGATCTCGTCTCATGCAACGATCTGGAAGCGCTTCCATGAGATCGGATCAGTTGC 198528
QY 1230 AGATGCTTTTCGGAAGAGTGTTGACAACTGACACACAGATATGGACCAAGGTGG 1289
Db 198529 GAGCGCTTTCGCGCGCTGGTGAAGCTGACACCTACCGCGGACATGGGCGCGCTGGCGCG 198588
QY 1290 CTACCTGGGACAGAGTGGCTCGACAGCTCACTATCGAGNAGCCCTGATACAGAGTT 1349
Db 198589 TTAATCTGGGCGGTTGGTTCCGAAGAGAGCTGGCTTGGCAGGAGCTCATCCGCGTGT 198648
QY 1350 AAGCCATCTCTTGTAGACGAAAGATATGAAGGCTTAAAAGGCTTAAAGCCAAATCTCGAATC 1409
Db 198649 CGATCATGCTGTATGACGACGAGATGCGGAGGCTCAAGGCTGAGATCTTGGCTTC 198708
QY 1410 GGGACTGAGCGTATGAGCGAGCTGTGAAGCAGCGAGTGGGCTCTGCACTACTATTAGAA 1469
Db 198709 GCGCTGTGCTGTGCGAGCTGGTTTCGACGCGGTGGGCTTGGCGCTTCCCTCCGCG 198768
QY 1470 CTTCGACAGCGCGGCGGTGCCAGGCTGACGATATAGCACTGGCGCCCAAAAAGACTG 1529
Db 198769 TTTGATGAGGCGGCGGCTGCGCAATGTCGCGGCGATCCGCGCTCAGCGCGGACGACTG 198828
QY 1530 GAGAGTTAAACACCTTCAGCACTTGGCAGGCTACTGAAACACTGAGAGATTCGCA 1589
Db 198829 GCGCGTCAACAGCGCGGCGGCTGCCAAGSGTGTGCAAGCTCCAGGCGATCGCA 198888
QY 1590 GGACTTTAAGCAGGCGCAATAGATTAACAAAGCAGTATGCTTTGGCGGACCTGATTTGCT 1649
Db 198889 GGACTTTAAGCGGCGGACAGCGGACGAGAGAGTCTGCTGTTGGCATCTGATGCTTCT 198948
QY 1650 GGGCGGCTGTGGGSGTGTAGAAAANGCTGCAAAAANGTGTGCCCTATGAGTGCAGGTGCC 1709

```

[illegible]

```

/translation="MDKSSAGKCPVAHTAPGRSNRQMPQLDVLVLRHSGLSD
PLENTAFETKEDDALKRDLRALITUSDPMPADFGHGGELFIMAWHSAGTYR
LGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
LEGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
NPGSNGNDPVAAADIRETFANMAHNDRETVALLAGHTEFKTHGAGGDFGADP
EGAGLEQGLGWGTFPGDAGTGGPEVTTNFKDAITGGPEVTTNFKDAITGGPEV
PAGATQWAKNAETITPDVADPSKHVPTITDLSRDFDAVEKISRRPLENPDFA
DAFAWELVSTHMSASSTPFGSKGGAGNARTLWQDVPAPDRBLRVDETDIAGLAKII
ASGLSVEGLTAMASSTPFGSKGGAGNARTLWQDVPAPDRBLRVDETDIAGLAKII
AGTQVMACTDGRKISGLDILAGANVAKKAGGHDITVPTTPORDASNAOTD
AGTQVMACTDGRKISGLDILAGANVAKKAGGHDITVPTTPORDASNAOTD
GPKVQVTSPEALNDFVNLNMGTONSPLEGGPEVGGESDGGRTGAAKGTGTVD
LIFGSHSGLAFAPATVAGSDAREKVKDFEVAATKVNNAAREFLV"
5640. 6113
/gene="Sma2381"
CDS
5640. 6113
/gene="Sma2381"
/function="Mifescellaneous; Unknown"
/translation="Glimmer prediction"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK65945.1"
/db_xref="GI:14524461"
/translation="MLKAMKIGQSILDTGDKWKLYCLAVRYSASETSAHALIR
TRANSLATION:MLKAMKIGQSILDTGDKWKLYCLAVRYSASETSAHALIR
IREKAGQSLVSLPARNVADFLVNLAIQINAHNGRHRHQLSLSKALAVY"
6167. 7135
/gene="Sma2383"
CDS
6167. 7135
/gene="Sma2383"
/function="Small molecule metabolism"
/translation="Glimmer prediction; similar to Tm22, a putative
note="Glimmer prediction; similar to Tm22, a putative
from cowpea, T11672. Also similar to the
oxidoreductases from Mycobacterium tuberculosis,
Agrobacterium tumefaciens, etc. Zinc-binding family"
/transl_table=11
/product="probable oxidoreductase"
/protein_id="AAK65947.1"
/db_xref="GI:14524462"
/translation="NKAVMKVGGVDYKVEVDEPVPARVAVVVEAAGVNFMDI
GVQGMADITIPNPKVLGVSGARVLVAGDGTGEAFVGDVRVANTVAPGSAQROSIPA
ASLVLPDVTDRDTPMKGALTAHEATDFVPVQGLDALVHAAGGVYGLLTQII
LRKGVNIGVSSDKVATMKGALTAHEATDFVPVQGLDALVHAAGGVYGLLTQII
RTPKGSIEALRSGTFCNPGVGLSGPELIMNLPKSIKIGTATHTRELLIDR
complement(7357..8946)
/gene="Sma2385"
CDS
complement(7357..8946)
/gene="Sma2385"
/function="Cell processes; Transport of small molecules"
/translation="Glimmer prediction; similar to probable ABC
transporter from Streptomyces coelicolor, T15245, and to
Streptomyces virginiae, BAA96297. ABC transporter family"
/codon_start=1
/transl_table=11
/product="probable ABC transporter, ATP-binding protein"
/protein_id="AAK65947.1"
/db_xref="GI:14524463"
/translation="STGSLKSWKPKDGHVPSDDIATCPHTGLVGRNGKETS
VIRASITSTGSLKSWKPKDGHVPSDDIATCPHTGLVGRNGKETS
SVLEETDMPTVRIYSALAHLEADPTLNLGSCQSTRAVAAATFSEPDII
DEPTNLRDGRRAVIGLLSGWRSGAIWVSHDRLELEMDAIIELTSLGTRYGGW
SAYQAAVALEAAQOGLTAKHTADEVDRKARALAEALRDKRDAISCTRAAKGDMRI
VLVGRKNSAESRGSVELAERAGADAVTAKARIEVLPQFIRLPTLELPAGRO
VLVPGTAGVDPARTIIRDLFSFLVGRPVSYCPNGSGKTSLLKLVYVGLPEPFKGT
VSNVYPTLLQGSVSLDERPTLEIKRNLKFMACRAALASPRADALQHVE
DEFLANIGLITGVELSVSG"
9057. 9713
/gene="Sma2387"
CDS
9057. 9713
/gene="Sma2387"
/function="Mifescellaneous; Unknown"
/translation="Glimmer prediction; putative bacterial regulatory
protein, terF family"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK65948.1"
/db_xref="GI:14524464"
/translation="ACSGSPSTREAEALPRLDLSSTFENKINGASRSP
IAEGLSIRLTYRTPRQALAAVWVEDLDFRFTKLAAGSGEAEALRQMAPR
YFRGRTDQDSAFYLSVTAGAVFSLPLRERLAWSAALPLVQATLSQAGLVPG
STLIECHVLIDLEGAANNRVRCLSESPASERLRLPESRMAVQPTAMHPNRPGLV"
complement(9922..10455)
/gene="Sma2389"
CDS
complement(9922..10455)
/gene="Sma2389"
/function="Cell processes; Adaptation"
/translation="Glimmer prediction; very similar to Ohr, a putative
stress-induced protein, from Bradyrhizobium japonicum,

```

```

Query Match 42.94; Score 960; DB 1; Length 10578;
Best Local Similarity 67.13; Pred. No. 4; Je-248;
Matches 1445; Conservativity 0; Mismatches 685; Indels 24; Gaps 5;
QY 85 CAAGTGTCCAGGTGGCGCACCAAAACAGGATTTGGTCCCAACATGCTCAACTCGGC 144
DB 5065 CATACGCGACGAGGACGAGATCGAACGCGACTGGTGGCGGATCAGTGGACGTCCAG 5006
QY 145 ATCTTACGCCCAATTCATGCTGTCGACCCCAACGACCGGATTTTGATATCCGGA 204
DB 5005 GTTCTCGGCGCATCTCGCGCTTTCGATCCGTGGGCANTAGCTCACTACCGCG 4946
QY 205 GAGTTTAAAGAACTAGATCGCGCGTTTAAAGAGACCTGGCAGCGCTAATGACAGAT 264
DB 4945 GAGTCAAGAAGACTCGACCTCGATCGGCTGAAGCGGACCTGCGTCACTGATGACGAT 4886
QY 265 TCACAGACTGGTGGCGACGAGATTCAGTGTGATATGGCCCTTCTTTATAGCGATGGCG 324
DB 4885 TCGCAGGACTCGTGGCGGCGGACTTGGCGGCTATGGCGGCTTTCNTCCGATGGCG 4826
QY 325 TGGCAGACGCGCGGACCTACCGTATCGGTGATGCGGCTGGTGGCGGCTCGCGGCTCA 384
DB 4825 TGGCAGACGCTGGCAGCTACCGTATCACCGATGCGCTGGCGGCGCGCGGCGGCTGAG 4766
QY 385 GAGTGTGGCGGCTGCTATGAGTGTGGCGGACGAGCGATCTGGANAGACGCTTG 444
DB 4765 CAGCGTTTCGCGCGCTCAATAGCTGGCGGACGAGCGCATCTCGACGAGCGCGACGT 4706
QY 445 CTCTTTTGGCGCATCAACAAAATACGCTCGAAAATCTCTCGGAGGACGACCAACCGGAT 504
DB 4705 TTGCTGTGGCGCATCAACAAAATACGCGGACGAGCATTTCTGGCGGATCTATGAT 4646
QY 505 CTGACAGCAAGACTAGCTGTGCGGACCTTTCGCGCTTTAAACTTTTGGTTCGAGTGGC 564
DB 4645 CTGACCGGCAATGTGCGGCTCGAATCCATCCGCTTCAAGACCTTCGTTTTCGCGGCG 4586
QY 565 AGACAGATGTATGGGACCTGAGAGAGATGTATATGGGAGGACGACCAACCGGATGGCTG 624
DB 4585 GGGGTGGACGATGTGAGGTACCGCGGACCGGACGACGACCTCGGCTG 4529
QY 625 GGAGCGAGCGGATGTGAGGTACCGCGGACCTGCGGACCTCGGAGCGGCTGACAGAT 684
DB 4528 GCGGAGGAGCGCTAGAGCGGACGAGCGCACTGAGCGAGCGCTTCGCGCGCTGCAAGT 4469
QY 685 GGACTCATCTATGTAACCCCGGAGGACCAAGCGGACGAGACCGCTATCGCTGCTGGG 744
DB 4468 GGTCTCATCTATGTAACCCCGGAGCGGACGAGCGGACCGGCTGCTGCTGGG 4409
QY 745 GGTGATTCGTGAGACTTTTGGCGGATGGCAATGANTGATGAGAGAAACCGTGGCTCTC 804

```







```
/locus_tag="PSP04439"  
/note="similar to SP.034520; identified by sequence  
similarity; putative."  
/codon_start=1  
/transl_table=11  
/product="ATP phosphoribosyltransferase"  
/protein_id="AAO57888.1"
```

Query Match	Score 951.2;	DB 1;	Length 311963;
Best Local Similarity	42.5%		
Best Local	55.5%		
Prod	1	10-2/5	

Best local similarity 95.0%;  
 Pseud. NO. 1.1E-243;  
 Matches 1471; Conservative 0; Mismatches 678; Indels 93; Gaps 2;

80 TTAAGCAAAGTCAGGTGGCGGCACCAAAACAGGGATTGGTGGCCCAACATGCTCAACC 139

b  
117501 TCACCCAGGTGCCGGGGGGGACACCAACCGTGACTGGTGGCCCAAGCAGCTGAATC 1175

140 TCGGCATCTTACGGCAACATTCATCGCTATCGGACCCAAACGACCCGGATTTTGACTATG 199

b 117561 TGAAGATTTTGACCAGCATTTCTTCGGCTCTCCGACCCGGATGGCGAGGACATTTGACATATG 11761

200 CCGAAGAGTTTAAGAGCTAGATCTGGCAGCGGTTAAAAAGGACCTGGCAGCGCTAATGA 259

b  
117621 CCAAGGAGTTC AAGAGCCTG TGAAGCGGTCAAGCAAGACCTGCGCGACGTAATGA 11766

260 CAGATTTCACAGGACTGGTGGCCAGCAGATTACGGTCATTATGGCCCCCTCTTTATACGCA 319

b  
117681 CCCGGTCGACGACTGGTGGCGGGCGGAATTTGGCCATACGGCCCGGTTCATTTCGA 1177

320 TGGCGTGGCCACAGCGCCGGCACCCTACCGTATCGGTGATGCGCTGTGTGGCGGTGGGCTCG 379

b

380 GATCTACAGCGGGTTCGCGCCGTTCATTATAGACTGCCGACGACCATGGCCCAATATATCCTCCCTTAATAAACAAC 420

[illegible][illegible][illegible][illegible]

500 TGAATCCTCAGAGGAACTGATGCTCGGAACTATGGGCTTAAACCTTTGGTATTTCGAC 559

117921 TCGTGTCTACCGGTAACAGTGTGGCGCTGGAGTCCATGGGCTTCAAGACCTTCGGGTTTTTCGG 117930

560 GTGGCAGAGCAGATGATGGAGCCTGAAGAAGATGTAATACTGGGGAGCAGAAACCGAAT 619

**b** 117981 GGGTCGCCCGGATGCTGGGAACCGGAAGACGCTGTACTGGGGTTCGGAACCACT 1180

620 GGCTGGGAGACAA-----532

b  
118041 GGCTGGTGGCGAAGAGCGGTACGGCGCGCAGAGAAATGCAGCAGCCGGGGCGACGGCA 1181

y y  
633 -----GGCTATGAGGTGACC 649  
| | | | |

b  
118101 CCCTGGTGGCGAGCCTGAGAACCCACGGCGAACGAAGAAGCCGCGACCGCGAGTGGCGGAC 118110

650 GAGAGCTCGAANAATCCCTGGGAGCCGTACAAATGGGACTCATCTATGTAAACCCCGAAG 709

b  
118161 GTAATCTGGAACCCGCTGGCCGCGTGAGATGGGCTGATCTATGTGAACCCGGAG 118222

710 GACCCAACGGCAAGCCAGACCCCTATCGCTGCTGCGCGTGATATTCGTGAGACTTTTGGCC 769

**b** 118221 GCCCTGAAGGCGTCCGGACCGTGTGCTTCGGCGAGGGATATCCGGAAACCTTCGGCC 118222

770 GAATGGCAATGAATGAAGAAACCGTGGTCTCATAGCGGGTGGACACACCTTCGGAA 829

**b** 118281 CGATGGCCATGAATGACGAGAGACCGTGGCGCTGATCGCGGTGGTCACGCCCTTCGGCA 1183

830 AAACCCATGGTGCTGCCGATCGGAGAAATATGTGGGCCGAGAGCCTGCCGCCCGCAGGTA 889

— — — — —

118341	AGACTCACGGTGGCCGCCGCCGACAC	---GTGGCTCAGAGCGCGGAACGACGGBTC	118393
990	TTGACAAATGAGCTCGGGGTGGAAAAACACT	ACGGCACCCGACACGGGTGGGATACCA	949
118398	TGACGACGAGCGCTTGGCTTGGACACAAAT	TTGGCACCCTGAAGGTGGCGACACA	118457
950	TCAGAGCTAGAGCGGTGGACACAGACCCCT	CTATGATGGAGAGCACTACTTTT	1009
118458	TCACAGTGGCTTGGAGTGACCTGGACCTG	CACTACCAATGAGCAGCACTGATACC	118517
1010	TTGAAAACCTCTTGGTTACAGGTGGAGCTT	TACAAAAGTACAGCTTGGAGCTTACGT	1069
118518	TGGAAACCTCTTTGGCTTTCGATGGGAAC	TGGCAAAAGCGCGCGCGCGCATCAGT	118577
1070	GGAAACCAAGAAGCGGTGGCGGGCTGGCA	CTATACCCGATGCAATGATCCCAAGCT	1129
118578	GGACGGCTAAAAAGCGCGTGGCGCTGGC	AGATTCGGATGCCAGCAGCCGACCAAC	118637
1130	CGCAGCTCCATATTGCTCTACTAGGACCT	GGCGCTGGCATCTGACCTGATTCAGMAA	1189
118638	GCATATTCAGAGCTACTATGAACACCTGA	CTGATGGCTGGCTTTCGCGAAGCATG	1249
118698	AGATTTCTCGCGCTTCTCGGCCACCCGAC	CACTGCGCGAGCTTGCAGCTGCTC	118757
1250	GGTACAACTGACACACAGAGATATGGACAT	GGCAAGTGGCTCTCTGGGACCAAGATGC	1309
118758	GGTTTCACTGTACTCACCGGACAGTGGCC	ACTCTGCGCGGTACTCGGCGCGGAACAC	118817
1310	CTCAGAGAGACTCATCTGGCAAGCCCTAT	TACCAGATGAGGCCATCTCTGTAGAGC	1369
118818	CGACCGAAGAGCTGTCTGCGCACGACCG	AGATCTCCAGACGCTCATCTCGCTGGGATG	118877
1370	AAAACGATATTAGGGCTTAAAGCGCAAA	TCTTGGATGGGACTGACGGTAAGCGACG	1429
118878	TGACAGTGTGGCGCTCAAGGCGAGATCT	TTTGAACCTTGGCTTTCGTGTACACAC	118937
1430	TGCTGACGAGCGCATGGCTTCTGATCAT	TTCTTGAACCTTGGCTTTCGTGTACACAC	1489
118938	TGGTTTCAACCCGATGGGGGCGAGTTC	CACTCTCGCGGCTCTGCACAAAGCGGCGG	118997
1490	CCACGGTGCAGTATACGACTGGCCCTC	ACAAAAGACTGGAGTAAACAACTCATG	1549
118998	CCATGGGGCTCTCGCTCTGTGGCGGAGAG	CTGTGGCGGTCAACCGCTTGACCT	13057
1550	AACTTGGCAGGGTACTCAAAACACTAGAG	GTATTCACAGAGACTTAAACAGCGCAAT	1609
119058	AACGTGGCAACGCTGTGATACGCTTGAA	GTATTACAGAGCAAGTCAACGCACTCAAT	119117
1610	CAGATACAAAGCGATTGCTTGTGGCCG	CACTGATCTGCTCGCGGTGTCCGGGTGAG	1669
119118	CAACGGCAGAGATGATATTCGCACTTGC	ATTCGACCTGCTGAGCGGCTAGTGGGCG	139177
1670	FAAAAGCTGAAAGATGCTGGCACTAGCTG	CAAGTGGCTCTTTCACCCGCGACAGCGG	1729
119178	ACGACGGCGGGAAGCAACCGCAGGTGAC	AGAGTAAAGGTTCTGATTCACAGAGGCG	119237
1730	ATGCCACCGCTGAGCAACCGATGGAGAG	CTTTTGAAGCACTAGAGCAGCGCTGACG	1789
119238	ATGGGTCTCAAGAGAGAGATGTGGTGTG	TTCTCTTGAGCCGATCGCGATG	139297
1790	GCTTTTGAAGAACTAATTAACCGGAGCAT	AAAGTATTCGCTGAGGAAAGTGTGTAGAC	1849
119298	GCTTTCGCACTATCAGAAAGGGCGTTA	CAAGATATCGCGCCAGATCGCTGTGTCACA	119357
1850	GGGCGCGCTTCTGCTCTTTGGGCAACGA	AAATGACTCTTGGTAGCGGTATGCGTG	1909
119358	TAGCGGTCTCAAGAGAGAGATGTGGTGT	GTCTGAGTCTGTGGCGGCTGCGGG	119417
1910	ACTTGGGCACTACAGCGGTGCGGACCAT	GTGCGAGTGTTCACAAATGACCGGTGAC	1969
119418	TCTTGAATATCAAGCTGGGCGACCAAGC	AGGTTGTACAGCAAGCGGAAACCC	119477

QY 1970 FATCCAGTACCTCTTTAAACCTGCTAGACCTCAACACTAAATGGCGAGCCAGCAT 2099  
 Db 119478 TGACCAAGCACTCTCTGAGAACCTGTTGAGCACTGGCTGTCGAGTGGAAAGCCG 119537  
 QY 2030 AATCAGCAAGTTTTTGAAGCAGAGACTTCAAAACTGGCGAAGTAAGTGAAGTGGCA 2089  
 Db 119538 GTGGCAAGCAAGCTTTGAAGCAGCAGTACCGCAAGCGGCGAAGTGAAGTGGACCGCT 119597  
 QY 2090 CGCGGGTGAAGCTTTGCGGATCAATTCGAGCTAGAGCCCTGGCGAGAGTGTACG 2149  
 Db 119598 CTGCTGACCTGCTGCTGGTTTCAACAGCTCAACTGCGTCCGATTCGAGGTGTACG 119657  
 QY 2150 GCTGTCAGATCTGACAGAAAGTTGTTAAAGATTTTGTGAAGCGCTGGCCCAAGTAA 2209  
 Db 119658 GCACCGCGATGCGCAGAAAGGTTTGTCAAGACTTCGTGGCGCTCTGGACCAAGTTAA 119717  
 QY 2210 TGGACCTGACGCGTTTGATCT 2231  
 Db 119718 TGAACCTGACCGTTTGGACT 119739

RESULT 8  
 SRFPS 2004 bp DNA linear BCT 05-MAR-1999  
 DEFINITION Streptomyces reticulii furs and cpeb genes.  
 ACCESSION V14317.1  
 VERSION V14317.1 GI:3758888  
 KEYWORDS catalase/oxidase; cpeb gene; Fe regulatory protein; furs gene.  
 SOURCE Streptomyces reticulii  
 ORGANISM Streptomyces reticulii  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomycetaceae.  
 AUTHORS Zou,P.J., Borovok,I., de Oude Lucana,D.O., Mueller,D. and Schrenpf,H.  
 REFERENCES The mycelium-associated Streptomyces reticulii catalase-peroxidase, its gene and regulation by Furs  
 Microbiol. 145, 549-559 (1999)  
 Borovok I. to 2504)  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JUL-1997) I. Borovok, Angewandte Genetik der Mikroorganismen, FB Biologie/Chemie, University of Osnabrueck, Barbarastrasse 11, Osnabrueck D-49069, FRG  
 REVIEWED revised by submitter 12-Oct-1998  
 RELATED entry: Y14336.1  
 FEATURES  
 1. Location/Qualifiers  
 1..2004  
 /organism="Streptomyces reticulii"  
 /mol\_type="genomic DNA"  
 /strain="ru45, described by H.Zahner, University of Tuebingen"  
 /db\_xref="taxon:1925"  
 /clone="posn"  
 /contig="11b-BamHI minilibrary"  
 78..533  
 /gene="furs"  
 78..82  
 /note="putative"  
 96..533  
 /note="furs"  
 /codon\_start=1  
 /evidence-experimental  
 /transl\_table=11  
 /product="Fe regulatory protein"  
 /protein\_id="CAA74697.1"  
 /db\_xref="GI:3758888"  
 /db\_xref="SPTREMBL:O87863"  
 /db\_xref="SPTREMBL:O87864"  
 GVAHVGHSLSQAVYALHAIAGLVRLIEPAGHPARYEKGWGNHIIHLYCRSGAR  
 ADVUCAVGAFCPLTAADHGHFAIDAEVYVYMGICPKXCTGSS"

gene

555..2864  
 /gene="cpeb"  
 555..560  
 /gene="cpeb"  
 567..3789  
 /gene="cpeb"  
 /EC\_number="1.11.1.6"  
 /codon\_start=1  
 /evidence-experimental  
 /transl\_table=11  
 /product="catalase/peroxidase"  
 567..3789  
 /db\_xref="GI:3758888"  
 /db\_xref="SPTREMBL:O87864"

terminator

2830..2864  
 /gene="cpeb"  
 /note="rho independent"  
 /note="rho 966 g 384 t  
 480 a 1074 c

BASE COUNT

480 a 1074 c 966 g 384 t

ORIGIN

Query Watch 41.78; Score 932.4; DB 1; Length 2904;  
 Best Local Similarity 65.98; Pred No. 1,2e-240;  
 Matches 1404; Conservative 0; Mismatches 716; Indels 12; Gaps 3;

QY

100 GCACCAAAACAGGAGTTGTGGCCCAACATGCTCAACCTCGCATCTTACGCCACAT 159

Db

663 GCGCGCGCACCCTGCTGGCGGAGCGCGCTCAACCTGAGATCTCCGCCAAGAC 722

QY

160 TCATCTCTATCGGACCAACAGCCGAGTTTGTACTATCGGAGAGTTTGAAGACTA 219

Db

723 CGGCGCTGGCCAAATCCGCTCGACAGGAGCTTGCATACGCGAGCCCTCAAGCCCTG 782

QY

220 GATCTCGCAGCTTAAAGAGACCTGGCAGCTATGACAGATTCACAGAGACTGTGG 279

Db

783 GACCTGCGCGCTGACGGGAGACATCGCGGCTGCTGACCATCGAGACTGTGG 842

QY

280 CCAGCAGATTACGCTCATTTATGCGCCCTCTTTATGAGCATGGGTGGCAGACGCCGC 339

Db

843 CGCGCGGACTCGCAACTACGCGCCGCTGATGATCGTATGGCTGGCAGACGCCGC 902

QY

340 ACCTACCTATCGCTGATGGCGCTGTGGCGGTGGCTCCGCTCACAGAGCTTCGCCCT 399

Db

903 ACCTACCTATCGCTGATGGCGCTGTGGCGGTGGCTCCGCTCACAGAGCTTCGCCCT 962

QY

400 CTCATATGTCGCCAGACATGCCAATCTGGATTAAGACAGCTTGTCTTGTGCCCATC 459

Db

963 CTCACAGCTGGCCGAGACAGCGCAACCTGGACAGCGCCGCTGCTGTCGCCGCTC 1022

QY

460 AACCAAAATAGCTGCAAAATCTCTCTGGCGGATCTATATATCTACACAGAAAGCTA 519

Db

1023 AAGAGAGTACGCCAGACATCTCTCTGGCGGCTGCTGCTCACGGGCAATGTC 1082

QY

520 GCTCTGCAAACTATGGGCTTTTAAATCTTTTGGTTTGGTGGTGGCAGACAGATGATG 579

Db

1083 GCCTGTGAGACAGTGGGCTTCAAGACCTTCGGCTTGGCGCGCGCTGGCGAGCTCGG 1142

QY

580 GAGCCTGAGACAGATGTATCTACTGTGGGAGACAGAACCGATGCTCTGGGAGACAGCGCTAT 639

1143 GAGCGGGAGGAGAGCTTACTGGGGTCCGAGAACCACTGCTGGAGAGACCGGGGGTAC 1202  
 640 GAAGGTGACGACAGCTCTAAATCCCTCGGGAGCGGTACAAATCGAGCACTATCTACTA 699  
 1203 ACCGGGACCGCGAGCTGGAGAACCCGCTCGGGCGGGTCCAGATGGGGCTGATCTAGTTC 1262  
 700 AACCCGGAAGACCCAAACGGCAAGCAGACCTATCGCTCGGGGGTATATCTGGTAG 759  
 1144 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT 1211  
 1263 AACCGGAGGGCTTCCAAACCGCAACCCGACCGATCGCGCGCGCGCGCGAGATCCGAGAG 1322  
 760 ACTTTTGGCCGATGAGCAATGATGAGAGAACCTGGCTCTCATAGCGGTGGACAC 819  
 1322 ACCTTCCCGCGCATGGCGATGATGACGAGAGAGACCTGCGCTGATCGCGCGGGCCAC 1382  
 820 ACTTCGGGAACCAACATGGTCTGGCGATGGCGAGAAATATGTGGGGCGAGGCGCTGCC 879  
 1145 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT 1212  
 1383 ACTTTCGAGGAGGAGGAGGAGGAGGCGGGCGG---ACCGTTCGGCGCCGACCGCG 1439  
 880 GCCCGAGGTATTGAAGAAATGAGCTGGGTGGGAAACCACTACGCGACGGACAGGT 939  
 1440 GCGCGTCTCCCTCGAGAGCAGCGCTGGGCTGGCGGACACCTACGCGACCGGCAAGGCG 1499  
 940 GGGATACCATCATCAGTGGAGTAGAGGGGCTGGACCAAGACCCCTACTCAATGGAGC 999  
 1500 GGGGAGGCTATCCCTCGGGCTGGAGTCACTTGGGACGACGGCGACCGAGTGGAGC 1559  
 1000 AATACCTTTTGAATACCTTTGTTGATGAGTGGGAGGTTACAAAGTTCACCTGGA 1059  
 1560 AACGGGTTCCTCAAGAACCTCTCGAGTACGAGTACGAGCTGGAGCAGACCGCGCGCGC 1619  
 1060 GTTATGATGGAAACCAAGACGGTGGCGGGCTGGACCAACCTCCGGATGCACATGAT 1119  
 1620 TCCACAGTGGTGGCGAGACGCCGGAG-----ATCATCCCGCGCGGCGAGC 1673  
 1120 CCCAGCAAGTCGCGACCTCAATATGCTCACTACGAGACCTGGGCTGGCATGGACCT 1733  
 1674 CGGTGGAAGAGACCGCGCGGAGTGTGACCAACCGACCTGTCCCTGGCGTTCGAGCGCG 1733  
 1180 GATTACGAATAATTTTCGACGGTACTATGAAACCCCTGATGATTTGCGATGCTTC 1239  
 1734 ATCTACGCGGATCTCCCGGTTCTACGAGAACCGGGAGGTTGCGGACCGCTTC 1793  
 1240 GCGAAGAGTGTCAACTGACACACAGAGATATGGGACCAAGGTGGCTACCTGGGA 1299  
 1794 GCGCGGCTGTGTGATGAGTCCACCACTGACATGGGCGGAGTGTGCTACTCGCGC 1853  
 1300 CCGAGGTCCTGAGGAGGCTGCTGGCGAGACCCCTATACAGAGTGAACCCCTCT 1359  
 1854 CCGGAGTCCCGGAGGAGACCTGCTGTGGGAGCCGCTGGGAGGTCGAGGCGGAGCG 1913  
 1360 CTTCGTAGACAAAGCATATTGAAGCGCTTAAAGCCAAAATCTCGGANTCGGATGAG 1419  
 1914 CTGATACAGACGCGCGACATCCGATCTCAAGCAACAGCTGTGGAGTTCGGCGCTGCC 1973  
 1420 GTAGCGAGTGTGAGGAGCGAGGGTCTGCATCTACTTTTAGAACTCTGAGAG 1479  
 1974 GTGTCCAGCTGCTACCACTGGGTGGCGCTCCCTACCTTCGCGCGGCGAGGAG 2033  
 1480 CCGCGGGTCCCAACCTGCACGTATGACACTGGCCGCCCAAAAAGACTGGGAATTAAC 1539  
 2034 CCGCGGGTCCGCAAGCGCCGATCTGCTGCGCCGCGAGCGGGCTGGGAGGTCAAC 2093  
 1594 AACCTCTACGATTCGAGGTGATCTCAAACTCAAGACTAGAGTATCCAGAGGACTTAAC 1599  
 2094 GACCCGACAGCTGGCGGAGGCTGGCGACCTCTGGAGAGCTCCAGCAGGATTCAC 2153  
 1600 CAGGCGCAATCAGATACAAAGACGATGCTGGTGGCGCACTGATTTGTCTGCGCGGCT 1659  
 2154 ---GGCTCTCGGGGGGAGAGGCTCTCCCTGCGCGGACCTCATGCTCTCGCGCGTCC 2210  
 1660 GCGGGTGTAGAAATGCTCAAAAGTCTGGCCATGAGTGTGAGTGTGAGTGTCAACCGG 1719  
 2211 GCGGGTGTAGAAAGCGCGCAAGGAGCGGGCTTGGAGTCCAGTGTGCTTCACCGG 2270

1720 GGAGGAGCGATGCCACCGCTGAGCAAAACGATGTGGAGACGCTTTCGAGACGCTAGAGCA 1779  
 2271 GSCCGTCTCCACGCGACCGAGAGAGCACTGAGTCTTCCGAGGCGCTTGGAGCGC 2330  
 1780 GCGGCTGACGGCTTTAGAAATCAATTAACCGGAGCATAAAGTATCGCTGAGGAATG 1839  
 2331 ACOCGAGAGCGCTCGCAACCTACTCGCAAGGGCAACCGCTGCGCGCGAGTACCTG 2390  
 1840 CTCTGAGAGCGGCGAGGCTGTCTGCTTGGCGAGCAAAAGTACGCTGTTGGTAGC 1899  
 2391 CTGCTCGAAGGCGAACTCTGAACTGAGCGCCCGAGATGACGCTGCTGCTCGT 2450  
 1900 GGTATGCTGCTAGTGGGCAACCACTAGACGTTTCGACGATGGAGTGTTCACAAATAG 1959  
 2451 GGTCTGCGGCTCTGGCGCCCAACCAACGACGCTGCGGCTTCCACAAAGC 2510  
 1960 CGGGGTGAGTATCAATCACTTTTGAACCTGTGAGCTCAAGACTAAATGGCA 2019  
 2511 CCGGGCTCTCAACCACTTCTGTCACTTCTGCTCACTGCTCGACATGGGACAGCTGGAG 2570  
 2020 GCCAGGATGAATCAGCAAAATTTTGAAGCGAGAGACTCAAACTGGCGAATTAAG 2079  
 2571 GCGACTGGGACGACGACGCTGAGGGCGGCGACGCCACCGCGGAGCTGGA 2630  
 2080 TGGAGTGGGACCGGGTACAGCTGATCTGCGATGCTGCGATGCTGAGAGCCCTGGA 2139  
 2631 TGGGCGGACCGGAGCGACCTGGTCTGCGCTCCAACTCGGAGCTGCGCGCTGCC 2690  
 2140 GAAGTGTAGGCTGTGAGTCTGAGAAAGTTTGAAGATTTTTCGAGAGGCTGG 2699  
 2691 GAGGCTACGCGCAAGCGAGCGAGAGAGTGTGTGAGGACTTCTGCGCGCGTGG 2750  
 2200 GCGAAGTATGAGCTGCGACCGGTTGATCT 2231  
 2751 CACAGSTCATGAGCGGCGGCTTGACCT 2782

RESULT 9  
 AY040244 4910 bp DNA linear BCT 03-OCT-2002  
 DEFINITION Burkholderia pseudomallei ATP-dependent DNA helicase protein RecG (recG) gene, partial cds; and oxidative stress regulatory protein OxyR (oxyR), catalase-peroxidase protein KatG (katG), and bpsA (bpsA) genes, complete cds.  
 ACCESSION AY040244 GI:22711992  
 VERSION 1  
 SOURCE Burkholderia pseudomallei  
 ORGANISM Burkholderia pseudomallei  
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Burkholderia; pseudomallei group.  
 REFERENCE 1 (bases 1 to 4910)  
 Loprasert,S., Salibhan,R., Whangsuk,W. and Mongkolsuk,S. The Burkholderia pseudomallei oxyR gene: expression analysis and mutant characterization. J. Microbiol. 140 (2002)  
 JOURNAL 2 (bases 1 to 4910)  
 REFERENCE 2 (bases 1 to 4910)  
 Loprasert,S., Salibhan,R., Whangsuk,W. and Mongkolsuk,S. Direct Submission  
 JOURNAL Submitted (14-JUN-2001) Lab. of Biotechnology, Chulabhorn Research Institute, Vipavadee, Bangkok 10210, Thailand  
 REFERENCE 3 (bases 1 to 4910)  
 Loprasert,S., Salibhan,R., Whangsuk,W. and Mongkolsuk,S. Submitted (07-JUN-2002) Lab. of Biotechnology, Chulabhorn Research Institute, Vipavadee, Bangkok 10210, Thailand  
 JOURNAL Sequence update by submitter  
 REFERENCE 4 (bases 1 to 4910)  
 Loprasert,S., Salibhan,R., Whangsuk,W. and Mongkolsuk,S. Direct Submission  
 JOURNAL Submitted (04-SEP-2002) Lab. of Biotechnology, Chulabhorn Research Institute, Vipavadee, Bangkok 10210, Thailand

REMARK  
COMMENT  
FEATURES

Sequence update by submitter  
On Sep 4, 2002 this sequence version replaced gi.22657796.

Location/Qualifiers  
1..4910

/organism="Burkholderia pseudomallei"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:28450"  
/contig="BPP0010000000"  
/clone="resG"  
/length=1183  
/gene="recG"  
/note="similar to Escherichia coli ATP-dependent DNA  
helicase recG c-terminal region"  
/codon\_start=1  
/trans\_table=11  
/protein\_id="AAK72464.1"  
/db\_xref="GI:16580119"  
/translation="EFTRGOSGAAMLRFLADLNDGLWLFEPADAAARLIANHPVVA  
QHLMWIGAREDTLKA"  
319..1278  
/gene="oxyR"  
/length=1572  
/note="similar to Escherichia coli OxyR"  
/clone="resG"  
/note="global regulator which belongs to the LysR family  
of transcription factors"  
/codon\_start=1  
/trans\_table=11  
/product="oxidative stress regulatory protein OxyR"  
/protein\_id="AAK72465.1"  
/db\_xref="taxon:28450"  
/translation="NTEELAYIVAVRHHGRRAAEACVFQPTLSVAIKLDELAL  
VQIFRGASEVSYTPGVQDQIVTQARVLEQTFATKIKAGQDPLGPRGLGVYIVTAL  
PVLPLVYKIVIRVYQPMPLMAENYTLKLLEKQGEIDAAINALPFPETGVLAVRAL  
YDEGFTVALLPAGHEWKKRKLDAADLKQETMLLGGKEDFRDVLGVCPDLAHFQSA  
XIDKFTVPSGSGVLTIRMYAGSGLTIVPMSVAEAGVGHASGPDADLLSYVPEDVPV  
DRWVWIKMSFTFRPAIDALISEATACGLPGVKLIDPATYN"  
4911..5000  
/gene="katG"  
/length=1704  
/note="similar to Escherichia coli catalase-peroxidase  
KatG"  
/codon\_start=1  
/trans\_table=11  
/product="catalase-peroxidase protein KatG"  
/protein\_id="AAK72466.3"  
/db\_xref="GI:22711993"  
/translation="MPSGDAQPRRRGVHEQRNRNNSAKCPFHQAAGNGTSNRDWWP  
NOLDLSIHRSSLSDEPMGQDFNYAQAFHLLDAVYKDPDLHALMTTSDWWPADFGH  
GLSTIRAMHSAFTYDTPRGAGGGEQGFAPLNSPWNANLAKRRLWPIKQYK  
RGLMADLITLGVNLYSNGGFTKTFGAGADYWPEDVYWGSEKILWELSGGPNR  
GAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGV  
TQSHNPNENLGGVEMELKSPAGAIQWAGVADNITPDAQDPKSHIPRMTLTDLS  
REFVAPLISREHNEKQFADANARAWFKPLTQVHMAAATFRGSDXRGAGANGRIILA  
PAVHPLIDADAELKALVLSGTLVSQVSTLWAAATFRGSDXRGAGANGRIILA  
PKQWNEQVPEOLAAVLETLERLRTAFNAGRGKGFVSLAGIYVLACAGVGEQAQKLL  
GHAVTVFAPRADAQSADEQTVSSNAVLPEADQGFNLSAGIYPAEVLVLDKQALL  
TILSPATVTVLGLGLVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGVAGV  
NLDSPATV  
4122..24910  
/gene="dpsA"  
/length=4422  
/note="non-specific DNA binding protein"  
/codon\_start=1  
/trans\_table=11  
/protein\_id="AAK04250.1"  
/db\_xref="GI:22657798"  
/translation="MAKNSNATQINIGSDKDRKKIANGLSLLADITVLYLTATNFH

gene  
CDS

gene  
CDS

gene  
CDS















QY	520	GCTCTGCAACTATCGCTTTAAAGCTTTGGTTTCGATTCGATGCGCAGACGAGATATGCG	579
DB	12889	GGCGTGGAGAGATGAGCTTTCGAGACCTTCGCGTTTCGCGCGCGCGCGAGCTGTGG	12830
QY	580	GAGCTCTGAAGAGATGTATATCTGGGAGCAGAAACGAATGGCTGGGAGACAGCGCTAT	639
DB	12829	GAGCGCGAGGAGAGCTCTACTTGGCGGCGAGACACATGGCTGCAGACCGCGCGCTAT	12770
QY	640	GAGGTGACCGGACGCTCGAATATCCCTGGGAGCGGTACGACATTCGACATGACATCA	699
DB	12769	ACCGGCGACCGGACGAGAGACCGCTCGCGCGCTCGAGATGGGTGTGATCTACGTC	12710
QY	700	AACCCGAAAGAGCAACACGCGACAGCCATCTGCTGCGCGCTGATATTCGTGAG	759
DB	12709	ACCGGAGGCGCCCAACAGCGACACCGCGGAGCGGATCGCGCGCGCGGACATCCGCG	12650
QY	760	ACTTTTGGCCGAATGCAATGATGACAGAAACGCTGCGCTCTATCCAGCGGTGACAC	819
DB	12649	ACCTTCCGCGCATGGCGATGACGACGAGAGACCGCTCGCGCTCATCGCGCGGACAC	12590
QY	820	ACCTTTCGGAAAAACCTGGTCTGCTCCGATGCGGAGAAATATGTGGCGGAGAGCTGCC	879
DB	12589	ACCTTTCGGCAAGACACGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	12533
QY	880	CGCGGAGATCTGAAGAAATGAGCGCTGGGAGAAACCTGACGACGACGACGAGCT	939
DB	12473	GCGCGCGCATGAGCAGCAGCGGCTCGGCTGGAAGAGCAGCCACGACGCGGAGGCG	12473
QY	940	GCGATATCATCACACCTGGATGAGAGCGGCTGCGACCAAGACCCCTACTCAATGAGC	999
DB	12472	GCGAGCGCTCATCTCGCGCGCGGAGTCTGAGACGACGACGACGACGACGACGACG	12413
QY	1000	AATACCTTTTGAAGAACTCTTTGTTAGAGTGGGAGCTTACCAAAAGTCGACGTGGA	1059
DB	12412	AACGGCTCTTCAAGAGACTCTTCGAGTTGAGTAGAGCTTGGAGCAGAGCGCGCGCG	12353
QY	1060	GCTTATAGTGGAAACAAAGAGCGTGGCGGCGTGGACCATACCGGATCGACATGAT	1119
DB	12352	GCGAACCTGTGGCGAGACGACCGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	12299
QY	1120	CCCAGCATGCGCAGCTCCATTATGCTGCTACGAGCTGGCGCTGGCGATGACGCTT	1179
DB	12298	CGCGCAGAGACGACCGCGCGCGGAGTGTCCACACGACGCTGTGCTGGCGCTGACCG	12239
QY	1180	GATTACGAAAAATTTCTGAGAGTACTATGAAACCTTGATGACTTTCAGATGCTTTC	1239
DB	12238	ATCTGCGCGCGATCTCCCGGCTGTACGAGACCGGAGGATTCGCGGAGCGCTTC	12179
QY	1240	GCGAAGCATGTACAACTGACACAGAGATATGGCAACAAAGTGGCTGCTACTCGGA	1299
DB	12178	GCGCGCGCTGGTTCAGCTGACCCAGCGACATGGCGGAGTCCCTGTACTCTGCGC	12119
QY	1300	CAGAGTGGCTCAGGAGACCTCATCTGCGAAGACCTTATCCAGATGTAGCATCCT	1359
DB	12218	CGGAGGAGGAGGAGCTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	12059
QY	1360	CTTGTAGCAGAAACGATATGAGGAGCTTAAAGCAAAATCTCGGATCGGAGTACG	1419
DB	12058	GTCATCGAGCGGAGGAGCTGCCACCGCTCAAGACAGCTCTCGAGTGGCGCTGTG	11999
QY	1420	GTMAGCGAGCTGGTACGAGGATGAGGCTTGTGATCTTCTTCTTCTTCTTCTTCTTCTT	1479
DB	11998	GTGTCGCGCTGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	11939
QY	1480	CGCGCGGCTGCGCAAGCTGACGATATGAGTGGCGGCGGACCAAAAGACTGGGAAGTAA	1539
DB	11938	CGCGCGGCTGCGCAAGCGCGCGGATCGGCTCGAGACCGAGCGCGGCTGGAGGTCAAC	11879
QY	1540	ACCTTCAGCTTTCGCGAGGATCTCAAAACACTAGAGGTATCCAGGAGGACTTTAAC	1599
DB	11878	GAGCGCGAGCTGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	11819

## RESULT 13

AE011760/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

da Silva, R.A.C.R., Ferro, J.A.,

Quaglini, R.B., Montello-Vi-

Camargo, L.E.A., Camarotte, G.,

Chambergo, F., Clapina, L.P.,

Cursino-Santos, J.R., E. Dorry,

Greggio, C.C., Gruber, A.,

Jr., R.P.P., Lemos, E.G.M.,

Madeira, A.M.B.N., Martines-

Menck, C.F.M., Miyaki, C.Y.,

Okura, V.K., Oliveira, M.C.,

Rossi, A., Sena, J.A.D., Silva,

Taticka, M., Temura, R.E.,

Santos, M., Trufelli, D.,

Tsal, S.M., White, F.F.,

Setubal, J.C. and

linear BCT 29-MAY-2002

section 138 of 469 of

the complete genome.

AE011760 AE008923

GI:21107450

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306

Xanthomonas axonopodis pv. citri str. 306



```

EVWTTTPAQNNDYDHLRGFEWELSKSPAGAHQWVAKNADATIPDHDASRKHERPT
LIDOLASDRAETKSLRSEIQQIPFADFAFARNTKLTIHNDGPRARLTGLADGVAE
EYDQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
NGARLRLAQEPKQNDPEOLAKVLAETLRLQADNAAAGSGKKLSLADLYLVACNAA
VHAQAQAQGVTPVPAAGTDAQSOBDYVESAVLRLPVADGEIRNFARRYAVFABAL
LIDAKOLLITAPLELVYLVGLVRLVGNVDFGSKGVPTSRPVLSNDFANLLDMRTE
WKATSEARVEYSGDRSETEGLRMTGTVDLVFGSINSILBAVEVYASADAQKEVHDF
VAAMTKVMDLQDFDLA*
complement(7505...8101)
CDS
complement(7505...8101)
/ gene="XAC1302"
/ note="Identified by sequence similarity; putative; ORF
located using Blastx/glimmer/Genemark"
/codon_start=1
/trans_table=1
/trans_table_id="transcribed hypothetical protein"
/db_xref="GI:21107458"
/translation="MTDOLAQGLQPOLQAGRLGAPQARSNVQAALPLMGALG
RNSQAGGTDAVLGLQDHAASPDLSGLSGLGSGNAGNGAVGHLFDKAP
QAAGLQGGTALVETSKASOLLQALPTVMYTLAKRFLSGEAASSSLQALPAQBHOV
RSPHAPGALLTVSDQDQGLGLDGLVYKLLGSLFGKR"
complement(8408...11116)
/ gene="XAC1303"
/ note="synonym: XAC1303"
complement(8408...11116)
/ gene="mutS"

```

Query Match 40.28; Score 900.2; DB 1; Length 11677;  
 Blast Local Similarity 64.98; Pred. No. 6.3e-232;  
 Matches 1437; Conservative 0; Mismatches 698; Indels 78; Gaps 4;

```

QY 94 GTGGCGGCACCAAAACAGGATTTGGGCGCCACATCGTCAACCTCGCATCTTACGC 153
Db 7013 GTTACCGGAACCAACATCGCATTTGGGCGCCACATCGTCAACCTCGCATCTTACGC 6954
QY 154 CMACTTCATCGTATCGGCACCAACGACCGCGATTTTGACTATCGGAGAGGTTTAA 213
Db 6953 CAGCATTTGACCAATCAACGACCGCGTGCACCGCTTCACATGCGCGCGCTTCAG 6894
QY 214 AAGCTAGATCTGGCAGCGGTTTAAAGACCTGGCAGCGCTAATGACAGATTCCAGGAC 273
Db 6893 CMCCTGGACCTACAGCGGCTCCACAGACGACCTCGCATGGTGTGATGACGATTCCGACGAC 6834
QY 274 TGGTGGCCACACATTAAGTGCATATATGTCGTCCTTTTATACGATGCGCTGGCACAC 333
Db 6833 TGGTGGCGCGCATTTGGCCATTTAGCGCGGCTTCTCGCATGGCTGGCTTAC 6774
QY 334 GCGGCACTACCGTATGCGGTATGCGCGGTGGTGGCGGTGGTGGCGGTGGTGGCGGTTC 393
Db 6773 GCGGCACTATGCGATTTGGGACGAGCGGTGGCGCGCGGACGCTGGGACGACGCTTC 6714
QY 394 GCGGCTCTCATATGTCGCGACACATGCGCAATCTGGATAGAGACGCTTCTCTTTG 453
Db 6714 GCGGCTCTCATATGTCGCGACACATGCGCAATCTGGATAGAGACGCTTCTCTTTG 453
QY 454 CCCATCAACAAATACGCTGCAAAATCTCTCTGGCGGATCTATGATTAATCAACAG 513
Db 6653 CCGATACAGAAATACGCGCAGGCGATTTCTGGGCGGATCTGATGATTTCTTACCGGC 6594
QY 514 ACSTAGCTCGMAACTATGGGCTTTAAACCTTTTGGATTTGCGGACGACAGAT 573
Db 6593 ACSTAGCTCGMAACTATGGGCTTTAAACCTTTTGGATTTGCGGACGACAGAT 573
QY 574 GTATTGGGACCTGAGAGATGTATATCTGGGACGAGAACCGATGGCTG---GGAGAC 630
Db 6533 ACCTGGGACCGGATACAGGATGTACTTGGGCGCGGAACCAAGTGGCTGGGTGTGAC 6474
QY 631 AAGCGCATAGAG----- 644
Db 6473 GAGCGCTATTTCGCGCGGCTTCGCGCGGCTGG8ACGACGCGCGGGGTGCTGGTGAAGGAC 6414

```

```

645 -----TGACCAGAGCTCGAAATCCCTGGAGCGGTACAAATG 684
Db 6413 GACGACACCCAGGTGGCGCATCCCGCGACCTGGAACACCGCTGGCGCGCTGGCAATG 6354
QY 685 GGACTCATCTATGTAACCCGACGACCTACGCGCAACCGACAGACCTATCGCTCGCTGG 744
Db 6353 GGCTCTATGTAGTCAATCGGAAGCGCCGAGCGCAACCCCGACCTCGATCGCTCGCGG 6294
QY 745 GGTGATNTCGTAGACTTTGGCCGAATGCAATGACGAGAAACCGTGGCTCTC 804
Db 6293 GCGGCAATCGGACACCTTTCGCGGCAATGCGCATGACGAGAAACCGTGGCGCTG 6234
QY 805 ATAGCGGCTGGACACCTTCGGAACCAACCTAGTGCTCGGATCGGAGGAATAATG 864
Db 6233 ATGCTTGGCGCGACACCTTCGCGCAAGACCAACCGCGCGCGCGCGCGAC---TACG 6177
QY 865 GCGCGAGAGCTTGGCGCGACGATGTGTGAAGAAATGACCTGGGCTGGAACCAACCTAC 924
Db 6176 GCGCGCAACCGGACGCGCGGCAAGCGGCTGGAAGCAAGCGGCTTCCGCTGGCAT 6117
QY 925 GCGACCGACGCTGGCGGCTTATCAGTGAACCAACCAACGACGCTGGCGGCTGGCACCA 984
Db 6116 GTTAGCGCACAGGCGCGCACCATCAGCGCGCTGGAAGTGAAGTGGACCAACCG 6057
QY 985 CTAATCAATGAGCAATTAATTTTTTGAACCTCTTGGTTAGAGTGGAGGTAC 1044
Db 6056 CTTGCGCAATGCGACGATTAATCGACCATCTCTCGGTTTCGATGGAGCTTAC 5997
QY 1045 AAAAGTCAGCTGAGCTTATCAGTGAACCAACCAACGACGCTGGCGGCTGGCACCA 1104
Db 5996 AAAAGTCAGCTGAGCTTATCAGTGAACCAACCAACGACGCTGGCGGCTGGCACCA 1104
QY 1105 CGGATGCAATGATCGACGACGACGCTGCTCATTTATGCTACTAGACGCTGGG 1164
Db 5942 CCGCGCGCGACGACGCTTCCGAGCAATGCGCGCGCGCTGACACCGATCTGGG 5883
QY 1165 CTGGCGATGGACCTGATTTAGCAAAAATTTCTCGACGCTACTTGAACACCTGATGAG 1224
Db 5882 CTGGCGCTTGGCGACGCGCTGACGACCATCTCGCGGCGCTTCCAGCATTCGCAAC 5823
QY 1225 TTTGCGAGTCTTTGGCAAGCGCTGACAAAGTGCACGACGACGACGACGACGACGAC 1284
Db 5822 TTTGCGGATGCTTGGCTGGCGCTGCTGACGCTCAGCTCAGCAGCATGGCGCGAC 5763
QY 1285 GTGCGCTACTTGGGACACGAAAGTGGCTTACGAGACGCTCATCTGGCAGACCTTACCA 1344
Db 5703 GAGCTACTTGGCGCTTGGCGACGCTTGGCGACGCTTGGCGACGCTTGGCGACGCTTGG 5703
QY 1345 GAGCTAGGCTCTTGTGACGACGACGACGATTTGAGGCTCTGAAGCGCAAAATCTG 1404
Db 5702 GCGCTGATCAGCACTTGTGATGCGACGACGCGCGCTGATTCGCTGGCGCGGACG 5643
QY 1405 GAATCGGAGCTGAGGTAAGGAGCTGGTGAACGACGATGGGCTTGTGCATCTACTTTT 1464
Db 5642 GCGCTGGCGCTTGGGCTTGGCGACGCTTGTGCTCATGCTGGGCTGGCGCTTGGCGCT 5583
QY 1465 AGAAGACTTGAACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1524
Db 5582 GCGGCTTGGCAACGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5523
QY 1525 GACTGGGAGTAACCAACCTCTACGACATTCGCAAGGCTACTCAAACTAGTAGAGTATC 1584
Db 5522 GAGTGGAGGCAATACGACCGCGACGACATTCGCGCAAGTGTGCGCAAGCTGGACGCTG 5463
QY 1585 CAGGAGACTTTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1644
Db 5463 CAGGCGGCTTTCAGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1644
QY 1645 GTGCTGGCGGCTTGGCGGCTGTAGAAAAGCTGCAAAAGTGTGGCCATGATGATGAC 1704
Db 5402 GTGCTGGCGGCTTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5343
QY 1705 GTGCTTTCAACCGCGGACGAGGAGTGGCGACGCGCTGAGCAACGATGTGAGAGCTTTC 1764

```

[illegible]









```

QY 1886 CTGCTTTGGTAGCGGTATGGGCTACCTAGGACCACTACGACGCTTCGACATGGAG 1945
Db 233437 CCGTGTGCTGGCGGTCTGGCGGTGCTCGGTGCCAATGCCAAGGTTCCGACACGGCG 233496
QY 1946 TGTTTACAAATAGCGGGTCAGCTATCCAAATGACTTCTTTGTAACCTGCTAGACCTCA 2005
Db 233497 TCCTCACCGGAAGGTGGGTGGCTGAGCAGAGACTTCTTCGTCAACCTGCTCGACATGA 233556
QY 2006 ACACATAATGCGAGCGACCGCATGANTCAGACAGTAGTGTGTAAGGACAGACTTCAAA 2065
Db 233557 GCAGCGCTGGGCGCC-----GGGCGCGAGCGCACCTACGAGGCTCGGACCGCAGAG 233610
QY 2066 CTGCGGAAGTAAAGTGGAGTGGCACCGGGGTAGACCTGATCTCGGATCCAAATTCGGAGC 2125
Db 233611 CCACCGGTGAAGTGGACCGCGGCGCGCGGATCTCATCTCGGCGGCACTCGCAGC 233670
QY 2126 TAAGAGCCCTCGCAGAGTGTAGCGGTGTGCAGATCTGTGAGAAAGTTCTTAAAGATT 2185
Db 233671 TCCGCGCTACGCCGAGGTCTATGCCACCTCGGACTCCAGGAGCACTTCGTGAGGACT 233730
QY 2186 TTGTGAAGGCTGGGCGCAAGTAATGGACCTGGACCGGTTTGATCT 2231
Db 233731 TCGCCAGGCGCTGGACCAAGGTGATGAACCTCGACCGGTTCCGATCT 233776

```

Search completed: October 8, 2003, 12:53:49  
 Job time : 8128.46 secs

Result	No.	Score	Query	Match	Length	DB	ID	Description
1	2238	100.0	2238	19	AAV06555			Microscilla furves
2	2238	100.0	2238	24	ABSS3945			DNA encoding Micro
3	844	37.7	2331	18	AAV04000			Myobacterium tube
C 1	844	37.7	2331	20	AAV00917			M. tuberculosis ca
C 2	844	37.7	2331	20	AAV00917			Myobacterium tube
C 3	844	37.7	443765	22	AAV09583			Myobacterium tube
C 4	844	37.7	4411529	22	AAV09583			Myobacterium tube
C 5	844	37.7	4411529	22	AAV09583			Myobacterium tube
C 6	844	37.7	2331	20	AAV00918			M. tuberculosis ca
C 7	844	37.7	2331	20	AAV00918			M. tuberculosis ca

XX NPI: 1998-085953/08.  
 XX DR P-F5DB, AMW3810.  
 XX PT New bacterial catalases, related nucleic acid vectors and  
 XX PT transformed cells - used as oxidising agents and for detecting or  
 XX PT destroying hydrogen peroxide, e.g. in biosensors  
 XX PS Claim 3; Fig 2; 35pp; English.  
 XX The present sequence is of the Micrococcus furvus catalase-33CAl  
 CC gene. Fragments of the gene can be used to identify related sequences.  
 CC Catalase-33CAl may be used to catalyse oxidation reactions such as  
 CC epoxidation or hydroxylation. The enzyme can also be used to detect or  
 CC destroy hydrogen peroxide, e.g. in connection with glycolic acid  
 CC production, biosensors, contact lens cleaning, pulp/paper bleaching and  
 CC pasteurisation of dairy products. Antibodies raised against  
 CC catalase-33CAl can be used to screen libraries for detection and  
 CC purification of cells containing the enzyme.  
 XX Sequence 2238 BP; 634 A; 545 G; 605 G; 454 T; 0 other;  
 SQ  
 Query Match 100.0%; Score 2238; DB 19; Length 2238;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 ATGGAATATCAACAACTCAGATCTTCTACGTATACACAACTATGCGGAAATGCG 60  
 Db 1 ATGGAATATCAACAACTCAGATCTTCTACGTATACACAACTATGCGGAAATGCG 60  
 61 CTTTTCACGAGGTGCGTTTATGCAAAAGTGTGCGGCGCCCAAAACAGGATGG 120  
 Qy 61 CTTTTCACGAGGTGCGTTTATGCAAAAGTGTGCGGCGCCCAAAACAGGATGG 120  
 Db 61 CTTTTCACGAGGTGCGTTTATGCAAAAGTGTGCGGCGCCCAAAACAGGATGG 120  
 121 TGGCCCAACATCTCAACTGCGCATCTTAGCCAACTATCATGCTATCGGACCAAC 180  
 Qy 121 TGGCCCAACATCTCAACTGCGCATCTTAGCCAACTATCATGCTATCGGACCAAC 180  
 Db 121 TGGCCCAACATCTCAACTGCGCATCTTAGCCAACTATCATGCTATCGGACCAAC 180  
 181 GACCGGATTTGACTATGCCGAGATTTAAGAGCTATGATCTGGCAGCGGTAAAGAG 240  
 Qy 181 GACCGGATTTGACTATGCCGAGATTTAAGAGCTATGATCTGGCAGCGGTAAAGAG 240  
 Db 181 GACCGGATTTGACTATGCCGAGATTTAAGAGCTATGATCTGGCAGCGGTAAAGAG 240  
 241 GACCTGGCAGCGCTATGACAGATTTACAGAGCTGTGGCCAGCAGATTTACGCTATT 300  
 Qy 241 GACCTGGCAGCGCTATGACAGATTTACAGAGCTGTGGCCAGCAGATTTACGCTATT 300  
 Db 241 GACCTGGCAGCGCTATGACAGATTTACAGAGCTGTGGCCAGCAGATTTACGCTATT 300  
 301 GCGCCCTCTTTATGCAATGCGCTGCGCATCTGCGGCGCTTCATAGCTATGCTGTGG 360  
 Qy 301 GCGCCCTCTTTATGCAATGCGCTGCGCATCTGCGGCGCTTCATAGCTATGCTGTGG 360  
 Db 301 GCGCCCTCTTTATGCAATGCGCTGCGCATCTGCGGCGCTTCATAGCTATGCTGTGG 360  
 361 GGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420  
 Qy 361 GGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420  
 Db 361 GGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420  
 421 GCAATCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
 Qy 421 GCAATCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
 Db 421 GCAATCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
 481 ATCTCTTGGGCGGATCTAATGATGATGATGATGATGATGATGATGATGATGATG 540  
 Qy 481 ATCTCTTGGGCGGATCTAATGATGATGATGATGATGATGATGATGATGATGATG 540  
 Db 481 ATCTCTTGGGCGGATCTAATGATGATGATGATGATGATGATGATGATGATGATG 540  
 541 AAATCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 600  
 Qy 541 AAATCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 600  
 Db 541 AAATCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 600  
 601 TGGGAGCAGCAACCAAGTGGTGGGAGCAGAGCTATGAAGGTGACCGAGAGCTCGAA 660  
 Qy 601 TGGGAGCAGCAACCAAGTGGTGGGAGCAGAGCTATGAAGGTGACCGAGAGCTCGAA 660  
 Db 601 TGGGAGCAGCAACCAAGTGGTGGGAGCAGAGCTATGAAGGTGACCGAGAGCTCGAA 660  
 661 ATGCGCGTGGGAGCGCTCAATATGCGGATCTATGATTAACCCCGAGAGCCAGCGG 720  
 Qy 661 ATGCGCGTGGGAGCGCTCAATATGCGGATCTATGATTAACCCCGAGAGCCAGCGG 720  
 Db 661 ATGCGCGTGGGAGCGCTCAATATGCGGATCTATGATTAACCCCGAGAGCCAGCGG 720

721 AAGCCAGACCTATGCTGCGCGGTATATCTGAGACTTTTGGCGAATGCAATG 780  
 Db 721 AAGCCAGACCTATGCTGCGCGGTATATCTGAGACTTTTGGCGAATGCAATG 780  
 781 AATGACGAAGAACCTGTGCTCTCATATGCGGTGACACACCTTGGGAACCCATGGT 840  
 Qy 781 AATGACGAAGAACCTGTGCTCTCATATGCGGTGACACACCTTGGGAACCCATGGT 840  
 Db 781 AATGACGAAGAACCTGTGCTCTCATATGCGGTGACACACCTTGGGAACCCATGGT 840  
 841 CTTCCGATCGGAGAAATATGTGGCGCGAGAGCTGCGCGCGCATGATTTGAAGAATG 900  
 Qy 841 CTTCCGATCGGAGAAATATGTGGCGCGAGAGCTGCGCGCGCATGATTTGAAGAATG 900  
 Db 841 CTTCCGATCGGAGAAATATGTGGCGCGAGAGCTGCGCGCGCATGATTTGAAGAATG 900  
 901 AGCTCGGGGTGAAACACCTTACGACCGCGAGCGCTGCGGATACATCAGCATGGA 960  
 Qy 901 AGCTCGGGGTGAAACACCTTACGACCGCGAGCGCTGCGGATACATCAGCATGGA 960  
 Db 901 AGCTCGGGGTGAAACACCTTACGACCGCGAGCGCTGCGGATACATCAGCATGGA 960  
 961 CTAGAAGCGCTTGACCAAGACCGCTTACTCAATGAGCAATACATTTTGAAGACCTC 1020  
 Qy 961 CTAGAAGCGCTTGACCAAGACCGCTTACTCAATGAGCAATACATTTTGAAGACCTC 1020  
 Db 961 CTAGAAGCGCTTGACCAAGACCGCTTACTCAATGAGCAATACATTTTGAAGACCTC 1020  
 1021 TTGTGGTTACGATGGAGCTTACCAAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAA 1080  
 Qy 1021 TTGTGGTTACGATGGAGCTTACCAAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAA 1080  
 Db 1021 TTGTGGTTACGATGGAGCTTACCAAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAA 1080  
 1081 GAGTGGTGGGCGGTGGCAATACCGATACCGATGACATGATCCCAAGGTGGACGCTCA 1140  
 Qy 1081 GAGTGGTGGGCGGTGGCAATACCGATACCGATGACATGATCCCAAGGTGGACGCTCA 1140  
 Db 1081 GAGTGGTGGGCGGTGGCAATACCGATACCGATGACATGATCCCAAGGTGGACGCTCA 1140  
 1141 TTATGCTCTACGAGACCTGGGCGTGGCATGAGCCCTGATACGCAAAAATTTCTCGA 1200  
 Qy 1141 TTATGCTCTACGAGACCTGGGCGTGGCATGAGCCCTGATACGCAAAAATTTCTCGA 1200  
 Db 1141 TTATGCTCTACGAGACCTGGGCGTGGCATGAGCCCTGATACGCAAAAATTTCTCGA 1200  
 1201 CGGTACTATGAAACCTGTGATGATTTGCAATGCTTTCCGCAAGCATGTTACAACTG 1260  
 Qy 1201 CGGTACTATGAAACCTGTGATGATTTGCAATGCTTTCCGCAAGCATGTTACAACTG 1260  
 Db 1201 CGGTACTATGAAACCTGTGATGATTTGCAATGCTTTCCGCAAGCATGTTACAACTG 1260  
 1261 ACACACAGATATGGGACCAAAAGTGGCTTCTGGGACAGAAAGTGGCTCAGGAAGAC 1320  
 Qy 1261 ACACACAGATATGGGACCAAAAGTGGCTTCTGGGACAGAAAGTGGCTCAGGAAGAC 1320  
 Db 1261 ACACACAGATATGGGACCAAAAGTGGCTTCTGGGACAGAAAGTGGCTCAGGAAGAC 1320  
 1321 CTGATCTGCGAGACCTATACAGATGTAAGCAATCTCTTGTAGAGAAACGATATT 1380  
 Qy 1321 CTGATCTGCGAGACCTATACAGATGTAAGCAATCTCTTGTAGAGAAACGATATT 1380  
 Db 1321 CTGATCTGCGAGACCTATACAGATGTAAGCAATCTCTTGTAGAGAAACGATATT 1380  
 1381 GAAGCGCTTAAAGAGCAAAATCTCGAATCGGAGTACGCTGAGCGAGCTGGTAAGCAG 1440  
 Qy 1381 GAAGCGCTTAAAGAGCAAAATCTCGAATCGGAGTACGCTGAGCGAGCTGGTAAGCAG 1440  
 Db 1381 GAAGCGCTTAAAGAGCAAAATCTCGAATCGGAGTACGCTGAGCGAGCTGGTAAGCAG 1440  
 1441 GCATGGGCTTCTGCACTTCTTTAGAACTCTGACAGCGCGGCTGGCGAGGTGCA 1500  
 Qy 1441 GCATGGGCTTCTGCACTTCTTTAGAACTCTGACAGCGCGGCTGGCGAGGTGCA 1500  
 Db 1441 GCATGGGCTTCTGCACTTCTTTAGAACTCTGACAGCGCGGCTGGCGAGGTGCA 1500  
 1501 GGTATACGATGGGCGCCCAAAAGATGGGAATTAACCAACCTCTCAGCACTTGGCCAGG 1560  
 Qy 1501 GGTATACGATGGGCGCCCAAAAGATGGGAATTAACCAACCTCTCAGCACTTGGCCAGG 1560  
 Db 1501 GGTATACGATGGGCGCCCAAAAGATGGGAATTAACCAACCTCTCAGCACTTGGCCAGG 1560  
 1561 GTACTCAAACTAGAGGTATCCAGAGGACTTTTACCAGGCGCATCAGATCAACAA 1620  
 Qy 1561 GTACTCAAACTAGAGGTATCCAGAGGACTTTTACCAGGCGCATCAGATCAACAA 1620  
 Db 1561 GTACTCAAACTAGAGGTATCCAGAGGACTTTTACCAGGCGCATCAGATCAACAA 1620  
 1621 GCAGTATGCTGGCGCACTGATGCTGCGCGGCTGCGGGGTGAGAAAAGCTGCA 1680  
 Qy 1621 GCAGTATGCTGGCGCACTGATGCTGCGCGGCTGCGGGGTGAGAAAAGCTGCA 1680  
 Db 1621 GCAGTATGCTGGCGCACTGATGCTGCGCGGCTGCGGGGTGAGAAAAGCTGCA 1680  
 1681 AAGATGCTGGCCTGAGGTGCAAGTGGCTTTCAACCGCGGAGCAGCGGATGCCACGCT 1740  
 Qy 1681 AAGATGCTGGCCTGAGGTGCAAGTGGCTTTCAACCGCGGAGCAGCGGATGCCACGCT 1740  
 Db 1681 AAGATGCTGGCCTGAGGTGCAAGTGGCTTTCAACCGCGGAGCAGCGGATGCCACGCT 1740  
 1741 GAGCAAAACGATGGAAGCTTTGGAAGACATAGAGCGCGCTGACGGCTTTAGAAAC 1800  
 Qy 1741 GAGCAAAACGATGGAAGCTTTGGAAGACATAGAGCGCGCTGACGGCTTTAGAAAC 1800  
 Db 1741 GAGCAAAACGATGGAAGCTTTGGAAGACATAGAGCGCGCTGACGGCTTTAGAAAC 1800



Db	601	TTGGGAGCAGAACCGATTGGCTGGGAGACAGGCGCTATGAGGTGACGGAGCTGCAA	660	1681	AAAGATGCTGGCCATGAGCTGCAGTGGCTTTCAACCCGGGACGAGCGGATGCCACGCT	1740
Qy	661	ATATCCCTCGGAGCGGTAAATGGGACTCATCTATGTAAACCCCGAGAGCCACGAGCG	720	1741	GAGCAACCCGATGGAGAGCTTTTCGAGACACTAGACCCAGCGCTTCGAGCTTTAGAAC	1800
Db	661	ATATCCCTCGGAGCGGTAAATGGGACTCATCTATGTAAACCCCGAGAGCCACGAGCG	720	1741	GAGCAACCCGATGGAGAGCTTTTCGAGACACTAGACCCAGCGCTTCGAGCTTTAGAAC	1800
Qy	721	AAGCCAGACCCATCTCGTCTGCTGGCGGTATATCTGAGAGACTTTTGGCCGAATGGCAATG	780	1801	TACATTTAAACCGGAGCATAAAGATTCGCTGAGGAAATTCGTAGACCGGCGGACGCTT	1860
Db	721	AAGCCAGACCCATCTCGTCTGCTGGCGGTATATCTGAGAGACTTTTGGCCGAATGGCAATG	780	1801	TACATTTAAACCGGAGCATAAAGATTCGCTGAGGAAATTCGTAGACCGGCGGACGCTT	1860
Qy	781	AATGACGAACCACTATGCTGCTGCGGCTGATATCTGTGAGAGACTTTTGGCCGAATGGCAATG	840	1861	CTGTGCTTTGGCCGACCAAGAAATGACCTTGGTGGGCGGTATGGGTACTTGGGCGAC	1920
Db	781	AATGACGAACCACTATGCTGCTGCGGCTGATATCTGTGAGAGACTTTTGGCCGAATGGCAATG	840	1861	CTGTGCTTTGGCCGACCAAGAAATGACCTTGGTGGGCGGTATGGGTACTTGGGCGAC	1920
Qy	841	CTGCGGAGTGGGGAATATGTGGGCGGAGAGCTGCGCCGACGATATGAGAATG	900	1921	AACTACGACCGGTTGGCAGCATGGAGTGTTCACAAATTAAGCCGGTCAAGTATCCAAATGAC	1980
Db	841	CTGCGGAGTGGGGAATATGTGGGCGGAGAGCTGCGCCGACGATATGAGAATG	900	1921	AACTACGACCGGTTGGCAGCATGGAGTGTTCACAAATTAAGCCGGTCAAGTATCCAAATGAC	1980
Qy	901	AGCCTGGGCTGGAAACACCTATCCGCGACGGACGCTGCGGATACATCACCAGTGA	960	1981	TTCTTTGTNAACCTGCTAGACTCAACCTCAACCTTAATGGCGACCCGACGATGATCAGACAA	2040
Db	901	AGCCTGGGCTGGAAACACCTATCCGCGACGGACGCTGCGGATACATCACCAGTGA	960	1981	TTCTTTGTNAACCTGCTAGACTCAACCTCAACCTTAATGGCGACCCGACGATGATCAGACAA	2040
Qy	961	CTAGAGGCGCTTGACACAGACCCCTACTCAATGGAGCAATTAATTTTGGAAACCTC	1020	2041	GTATTTTGAAGCGACAGACTTCAAACTGGCGAGTAAAGTGGAGTGGCACCCGGGTAGAC	2100
Db	961	CTAGAGGCGCTTGACACAGACCCCTACTCAATGGAGCAATTAATTTTGGAAACCTC	1020	2041	GTATTTTGAAGCGACAGACTTCAAACTGGCGAGTAAAGTGGAGTGGCACCCGGGTAGAC	2100
Qy	1021	TTTGCTTACAGTGGGAGCTTACAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAA	1080	2101	CTGATCTTCGGATCCATTCGAGCTAGTGAAGCCGCTGCGAGGATGAGCGGTGCGAGT	2160
Db	1021	TTTGCTTACAGTGGGAGCTTACAAAGTCCAGCTGAGCTTATCAGTGGAAACCAAA	1080	2101	CTGATCTTCGGATCCATTCGAGCTAGTGAAGCCGCTGCGAGGATGAGCGGTGCGAGT	2160
Qy	1081	GAGCGTCCGCGGCTGGCAATACCGATATGACATGATCCACAGCAAGTGGCGATCCCA	1140	2161	TCTGAAGAAAAGTTTGTAAAGATTTTGTGAAGCGCTTGGGCCAAAGTAAATGACCTGGAC	2220
Db	1081	GAGCGTCCGCGGCTGGCAATACCGATATGACATGATCCACAGCAAGTGGCGATCCCA	1140	2161	TCTGAAGAAAAGTTTGTAAAGATTTTGTGAAGCGCTTGGGCCAAAGTAAATGACCTGGAC	2220
Qy	1141	TTATGCTCTACTACGAGGCTGGCGGCTGGCGATGGACCTGATTAACGAAAAATTTCTCGA	1200	2221	CGGTTTGTATCTGAAATAA	2238
Db	1141	TTATGCTCTACTACGAGGCTGGCGGCTGGCGATGGACCTGATTAACGAAAAATTTCTCGA	1200	2221	CGGTTTGTATCTGAAATAA	2238
Qy	1201	CGGTACTACGAGGCTGGCGGCTGGCGATGGACCTGATTAACGAAAAATTTCTCGA	1260	RESULT 3		
Db	1201	CGGTACTACGAGGCTGGCGGCTGGCGATGGACCTGATTAACGAAAAATTTCTCGA	1260	NT90400		
Qy	1261	ACACAGAGATATGGGACCAAAAGTGCCTACTCTGGACACAGAGTGCCTCAGGAAGC	1320	ID	AAT90400 standard; DNA; 2331 BP.	
Db	1261	ACACAGAGATATGGGACCAAAAGTGCCTACTCTGGACACAGAGTGCCTCAGGAAGC	1320	AC	AAT90400;	
Qy	1321	CTCATCTGGCAGACCTTATACAGATGTAGGCAATCTCTCTGTAGAGCAAAAGATTT	1380	XX	25-MAR-2003 (updated)	
Db	1321	CTCATCTGGCAGACCTTATACAGATGTAGGCAATCTCTCTGTAGAGCAAAAGATTT	1380	DF	21-JAN-1998 (first entry)	
Qy	1381	GTGAGGAGGCTTATGAGAGTATCCAGAGGCTTAAAGCGGATGAGGATGAGGATGAGG	1440	DE	Mycobacterium tuberculosis wild-type katG gene.	
Db	1381	GTGAGGAGGCTTATGAGAGTATCCAGAGGCTTAAAGCGGATGAGGATGAGGATGAGG	1440	KW	Isoniazid resistance; restriction fragment length polymorphism;	
Qy	1441	GCATGGGCTTCTGCACTACTTTTGAAGAACTCTGACAGCGCGCGGTGCCAAGCTGCA	1500	KW	REP; katG gene; catalase; peroxidase; multiple drug resistance;	
Db	1441	GCATGGGCTTCTGCACTACTTTTGAAGAACTCTGACAGCGCGCGGTGCCAAGCTGCA	1500	KW	isonicotinic acid hydrazide; INH; tuberculosis; antibiotic;	
Qy	1501	CGTATACGCTGGGCCCCAAAAAGACTGGGAAGTAACACCTCAGCAACTTGGCCAGG	1560	XX	Mycobacterium tuberculosis.	
Db	1501	CGTATACGCTGGGCCCCAAAAAGACTGGGAAGTAACACCTCAGCAACTTGGCCAGG	1560	PH	Key Location/Qualifiers	
Qy	1561	GTACTCAAAACACTTGAAGTATCCAGAGGCTTAAAGCGGATGAGGATGAGGATGAGG	1620	FT	CDS	
Db	1561	GTACTCAAAACACTTGAAGTATCCAGAGGCTTAAAGCGGATGAGGATGAGGATGAGG	1620	FT	70..2292	
Qy	1621	CGATGATCTGTGGCCGACGATGATGCTGCGCGGCTGTGCGGTGTGAAAAAGCTGCA	1680	FT	a	
Db	1621	CGATGATCTGTGGCCGACGATGATGCTGCGCGGCTGTGCGGTGTGAAAAAGCTGCA	1680	FT	859	
Qy	1681	AAAGATCTGGCCATGAGCTGACGCTGCTTTCACCCGGGACCGGATGCCACGCT	1740	FT	b	
Db	1681	AAAGATCTGGCCATGAGCTGACGCTGCTTTCACCCGGGACCGGATGCCACGCT	1740	FT	/note- *G to C transversion at base 959 alters codon 264 from GCG (Ala) to ACC (Thr)	
Qy				FT	/note- *G to C transversion at base 1013 alters codon 315 from AGC (Ser) to ACC (Thr) leading to isoniazid resistance, and produces a new HspI restriction site.	
Db				FT	1013	
Qy				FT	/note- *G to C transversion at base 1013 alters codon 315 from AGC (Ser) to ACC (Thr) leading to isoniazid resistance, and produces a new HspI restriction site.	
Db				FT	1079	
Qy				FT	mutation	





QY 1410 GGAGCTACAGCTTAAAGCAGCTGGTAAGCAGCGATCGGCTCTCGACTACTCTTTAGAA 1469  
 Db 1460 TTTTCTTCTCTCAGAGCTTCTGAGCGGCTGGGCGGCGCTGCTGCTCGCG 1523  
 QY 1470 GTCTGACAGCGGGGCTGCAAGCGTGCAGCTATAGCACTGCGCCCAAGAAAGCTG 1529  
 Db 1520 TTTTCTTCTCTCAGAGCTTCTGAGCGGCTGGGCGGCGCTGCTGCTCGCG 1583  
 QY 1530 GGAGCTAAAGCAAGCTCA---GCAACTTCCGACGGTACTCAAAAGCTAGAGGTATCCA 1586  
 Db 1580 GAGCTTCAAGCGCCGAGCGGATCTCCGACAGGCTATTCGACCTCTGGAGAGATCCA 1643  
 QY 1587 GGAGGACTTTTAAACGACGCAATCAGATCAAGCAAGCTATGTTGGCGAGCTGATGT 1646  
 Db 1640 TTTTCTTCTCTCAGAGCTTCTGAGCGGCTGGGCGGCGCTGCTGCGGCAAGCTGCTGT 1703  
 QY 1647 GTGCGCGGCTGCGGCTGTAGAAAGCTGCAAAAGATGCTGCCATGAGGTGCAAGT 1706  
 Db 1700 GCTCGTGGCTTCTGCGGCTATGAGAAAGCAAGCGGCTGGGCGGCTGCAAGCTGAGT 1766  
 QY 1707 GCTCTTCAACCGGACGAGCGGATGCAACCGTGAAGCAAGCTGTTGGAAGCTTTGCA 1766  
 Db 1760 GCGCTTCAACCGGACGAGCGGATGCAACCGTGAAGCAAGCTGTTGGAAGCTTTGCA 1823  
 QY 1767 AGCAGTAGAGCAGCGCTGAGCGGCTTGAAGACTATTAAGCGGAGCAATGAAGTATC 1826  
 Db 1820 GCTGCTGAGCGGAGCGGATGCTTCCGAAGCTACTGCGAAGCGCAAGCGGCTTGC 1883  
 QY 1827 GCTGAGGAAATGCTGCTGAGCGGCGGAGCTTCTGCTGCTTTCGGCACCAAGATGAC 1886  
 Db 1880 TTTTCTTCTCTCAGAGCTTCTGAGCGGCTGGGCGGCGCTGCTGCTGCTGCTGCTG 1943  
 QY 1887 GCTGTTGGTGGGCTGCTGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGT 1946  
 Db 1940 TTTTCTTCTCTCAGAGCTTCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGT 2003  
 QY 1947 GTTTTCAAAATGAGCGGCTGAGCTATCCAAATGACTTCTTGAACGCTGCTAGACTCAA 2006  
 Db 2000 TTTTCTTCTCTCAGAGCTTCTGAGCGGCTGGGCGGCGCTGCTGCTGCTGCTGCTG 2063  
 QY 2007 CACTAAATGCGGACGCGGAGTACGACAAAGCTTCTGAGCAAGCTGCTGCAATGGG 2066  
 Db 2060 TATCAGCTGGGAGCGCTGCGGACGAGATGACGCGGAGCTACAGGCGAGGATGCG---AG 2120  
 QY 2067 TGCGAGTAAAGTGAAGTGGCAAGCGGCTGAGCTGATCTTGGGATCCAAATTCGAGCT 2126  
 Db 2120 TGGCAGGTGAAGTGGGCTGAGCGGCTGAGCTGATCTTGGGATCCAAATTCGAGCT 2180  
 QY 2127 AAGGAGCTCGCAGAGTGAAGCTGCTGCAAGCTTCTGAGCAAGCTTCTGTAAGGTT 2186  
 Db 2180 GCGGCGCTTGTGCAAGCTTATGCGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTTCTG 2240  
 QY 2187 TGTGAAGGCTCGGCGCAAGTATGAGCACTGGACGGGTTGATCTG 2242  
 Db 2240 GTCGCTGCTGGGACAGGATGATGACCTGCAAGGCTGCAAGCTG 2286

RESULT 4

AAAX00817

ID MAX00817 standard; DNA: 2331 BP.

AC AAX00817;

XX 26-MAR-1999 (first entry)

XX M. tuberculosis catalase peroxide (katG) gene sequence.

DE Catalase-peroxide; katG; isonicotinic acid hydrazide; tuberculosis;

KW INH; diagnosis; detection; ss.

QY Mycobacterium tuberculosis.

XX

QY

PH

Key Location/Qualifiers

FT CDS 70..2292  
 FT /\*tag= a  
 FT /gene= katG  
 FT /product= catalase peroxide\*  
 FT /note= "the start codon is not indicated"  
 XX  
 PN M09850585-A1.  
 XX  
 PD 12-NOV-1998.  
 XX  
 XX 06-MAY-1998; 98NO-US09285.  
 XX  
 PR 07-MAY-1997; 97US-0852219.  
 XX  
 XX (MAYO-) MAYO FOUNDATION.  
 XX  
 XX Cockerill FR, Kline BC, Uhl JR;  
 XX  
 XX 97J, 1999-070899/06.  
 DR  
 DR P-FDB; M095398.  
 XX  
 PT Detection of Mycobacterium tuberculosis - by amplifying katG gene  
 PT and detecting specific fragment, and optionally identifying  
 PT INH-resistant strains by detecting specific mutation  
 XX  
 XX Claim 3; Fig 7; 83pp; English.  
 XX  
 CC The invention relates to a novel method of detecting Mycobacterium  
 CC tuberculosis. The method comprises amplifying the DNA in the samples to  
 CC generate a detectable amount of amplified DNA comprising a catalase-  
 CC peroxide (katG) DNA fragment with sequence of bases 904-1523 of the  
 CC M. tuberculosis strain H37Rv MC katG gene sequence. The method optionally  
 CC further comprises determining the presence of a specific mutation in an  
 CC isonicotinic acid hydrazide (INH)-resistant M. tuberculosis strain. The  
 CC method can be used to detect M. tuberculosis in biological fluids. The  
 CC especially human sputum, useful to diagnose tuberculosis. This disease is  
 CC a major cause of human morbidity and mortality, and conclusive diagnosis  
 CC and subsequent treatment depends on identification of the etiologic agent  
 CC M. tuberculosis. INH has been used in tuberculosis treatment, but  
 CC INH-resistant strains have emerged; the method allows such drug-resistant  
 CC strains to be detected. The method represents the wild-type  
 CC M. tuberculosis katG gene sequence.  
 XX  
 SQ Sequence 2331 BP; 455 A; 734 C; 763 G; 379 T; 0 other;  
 Query Match 37.7%; Score 844.4; DB 20; Length 2331;  
 Best Local Similarity 63.0%; Pred. No. 1.8e-253;  
 Matches 1390; Conservative 0; Mismatches 801; Indels 15; Gaps 5;  
 QY 30 TAGCTATACACCAACCTGCGGCAAAATGCGCTTTACGCGAGCTGCGTTAGCAAG 89  
 Db 93 TACAGAAACCAACCGGAGCGGCTGACCAAGCGCTGCTCGGCTCATATGAATA 152  
 QY 90 TGCAGGTGGCGCACCAACCAAGGAGTGTGGCCCAACATGTCACCTGGCATCTT 149  
 Db 153 CCGGCTGAGGCGGCGGAGACAGACTGTGGCCCAACCGCTCATCTGAGGACT 212  
 QY 150 AGCCACATCATCTCTATGCGACCCCAACGACCGGATTTTCATATGCGGAGGTT 209  
 Db 213 GCACCAACCAACCGGCGGCTGCTGACCGGCTGCGGCGGCTTTCGACTATGCGGAGGT 272  
 QY 210 TAAGAACTAGATCTGGCAAGCGGTTAAAGAGCACTGGCAGCGCTAATGACAGANTCA 269  
 Db 273 CGGCACTAGCTGTCACCGCTGCGCGGCGGACATCGAGAGTGTGACCACTCGCA 332  
 QY 270 GAGCTGTGGCGCAGATAGCGTCAATATGCGCCCTCTTTATACGATGCGGCGCA 329  
 Db 333 GCGCTGTGTGGCGCGCGGCTGCTGATCGGCGGCTGCTTATCCGAGTGGCTGCA 392  
 QY 330 CAGCGCGCGGCACTACCTATCGGTATGCGGCGGCTGCTGCGGCTGCGGCTGCAAGG 389  
 Db 393 GCTGTGCGGCACTACCGCATTCAGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 452











Qy 2127 AAGAGCCCTGCCAGACTTACGCGTGTGCAGATCTTGAAGAAAGTTTGAAGATTT 2186  
 Db 2153958 GCGGGCGCTTGTGAGCTCTATGGCGCGATGAGCGGAGTCTGTGCGAGACTT 2153939

Qy 2187 TGTGAAGAGCTGGGCGCAAGTAATGGACCTTGGACCGCTTTTGAATCTG 2232  
 Db 2153938 CGTCGCTGCTGGGCAAGGTGATGACCTTCGACAGGTGCGACGTG 2153893

RESULT 8  
 AAX00818  
 ID AAX00818 standard; DNA; 2331 bp.  
 AC AAX00818;  
 AT AAX00818;  
 DT 29-MAY-1999 (first entry)  
 XX M. tuberculosis catalase peroxide (katG) gene variant.  
 DE  
 XX Catalase-peroxide; katG; mutation: Isonicotinic acid hydrazide;  
 KW INH; tuberculosis; diagnosis; detection; variant; as.  
 XX  
 XX Mycobacterium tuberculosis.  
 OS Synthetic.  
 FH Location/Qualifiers  
 FT Key 70..2292  
 FT CDS /\*tag= a  
 FT /\*product= "catalase peroxide variant"  
 FT /\*note= "the start codon is not indicated"  
 FT 1013  
 FT 1013  
 FT /\*note= "g to C variation as compared to the  
 FT wild-type katG gene"  
 XX W09850585-A1.  
 XX 12-NOV-1998.  
 XX 06-MAY-1998; 98W0-US09285.  
 XX 07-MAY-1997; 97US-0852219.  
 XX (MAYO-) MAYO FOUNDATION.  
 XX Cockrell FR, Kline BC, Uhl JR;  
 XX WPI: 1999-070099/06.  
 XX P-PSDB: AAM95399.  
 PT Detection of Mycobacterium tuberculosis - by amplifying katG gene  
 PT and detecting specific fragment, and optionally identifying  
 PT INH-resistant strains by detecting specific mutation  
 PS Claim 3; Page -; 83pp; English.  
 XX  
 XX The invention relates to a novel method of detecting Mycobacterium  
 CC tuberculosis. The method comprises amplifying the DNA in the samples to  
 CC generate a detectable amount of amplified DNA comprising a catalase-  
 CC peroxide (katG) DNA fragment with sequence of bases 904-1523 of the  
 CC Mycobacterium strain H37Rv. The method optionally  
 CC further comprises amplifying a second DNA fragment with a serine to  
 CC threonine mutation in codon 315 (S315T mutation) of the Mycobacterium  
 CC Isonicotinic acid hydrazide (INH)-resistant M. tuberculosis strain. The  
 CC method can be used to detect M. tuberculosis in biological fluids,  
 CC especially human sputum, useful to diagnose tuberculosis. This disease is  
 CC a major cause of human morbidity and mortality, and conclusive diagnosis  
 CC and subsequent treatment depends on identification of the etiologic agent  
 CC Mycobacterium. INH has been used in tuberculosis treatment, but  
 CC INH-resistant M. tuberculosis strains have been identified. INH-resistant  
 CC strains to be identified. The present sequence represents the H.  
 CC tuberculosis katG gene variant.

CC Note: this sequence is not provided in the specification; it has been  
 CC created by modifying the katG gene sequence given in Fig 7.

Qy Sequence 2331 bp; 455 A; 735 C; 762 G; 379 T; 0 other;

Query Match 37.7%; Score 842.8; DB 20; Length 2331;  
 Best Local Similarity 63.0%; Pred. No. 5.8e-253;  
 Matches 1389; Conservative 0; Mismatches 802; Indels 15; Gaps 5;

Qy 30 TACGTATACACAAACATCGGCGAAMATGCCCTTTTACGGAGTGTTCGCTTAAGCAAG 89  
 Db 93 TACAGAGTACACACCCGAGCGCGCTAGCAACGGCTGTCCGCTGTGGGTTCATATGAATTA 152  
 Qy 90 TCGAGGTGGCGCGACCAAAACAGAGATGTGGTGGCCACATGCTCAACCTCGGCACTTT 149  
 Db 153 CCCCCTCGAGGGGGCGGAAACACAGAGCTGTGGCCCAACCGGCTCAATCTGAAGGTACT 212  
 Qy 150 AGCGCAACATTCATGGCTATGCGACCAACACCGGAGTTTGACTATGCGCAAGATT 209  
 Db 213 GCACCAAAACCGGCGCTGTGATCCGATGAGTGGCGCTTGCATATGTCGCGAGGT 272  
 Qy 210 TAAAGAGCTAGATCTGCGAGCGGTATAAAGAGACCTGGCAGCGCTTAATGACAGATTACA 269  
 Db 273 CCGCAGCAATGAGCTTGAAGCCCTCGACCGGCACATCGAGAAAGTATGATGACCACTCGCA 332  
 Qy 270 GCAATGGTGGCCACACATTAAGTGTATATGCCCCCTTCTTTATAGCGATGGCGTGGCA 329  
 Db 333 GCGGTGTGGCGCGCCACATACGGCCATACGGCCGCTGTATTCGGATATGGCTATGGCA 392  
 Qy 330 CAGCGCGCGGACCTACCGCTATCGGTATGCGCGTGTGGGCGTGGCTCCGGCTCAAGCG 389  
 Db 393 CGCTGCGGCACACTACCGCATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 452  
 Qy 390 CTTGGCGCGCTCATAGCTGGCGACAGATGCCCATCTTGGATAGACACCTTGGTCTT 449  
 Db 453 GTTCGCGCGCTTTAAACAGCTGGCGCGACACGCCACCTTGGACAGCGCCCGCGCTGCT 512  
 Qy 450 TTGGCCCATCAACAAAAATACGGTGGAAATCTCTCTGGGCGGATCTAATGATACTAC 509  
 Db 513 GTGGCGGCTCAGAAGAAAGTACGGCAAGAGCTCTCATGGGCGACCTGATGTGTTGCG 572  
 Qy 510 AGCAACCTAGCTCTGGCAAGTATGGCGTTAAACATTTGCTTCAGTTCAGTTCAGTTCAG 569  
 Db 573 CGCGCACTGCGCGCTTGAATCATGGCTTCAAGAGCTTCGGGTTCGGCTGGCGCGGT 632  
 Qy 570 AGATGTATGGAGCTCGAAGATGATATCTATGGGAGCAGAAACCGAATGGCTGGGAGA 629  
 Db 633 CGCCAGTGGGAGCC--CGATGAGTCTATTTGGGCGAAGAGACCACTGGCTGGCGCA 689  
 Qy 630 CAAGCGCTATGAAGCTGCGGAGCTCGAATAATCCCTGGGAGCGCTTACAAATAGGACT 689  
 Db 690 TGAAGCTTACAGCGGTAAAGCGGATCTGGAGAACCGGCTGGCGCGCTCAGATGGGGCT 749  
 Qy 690 CATCTATGTAAACCGAAGGACCCCAACGGCAGCCCTATCGCTGCTGGCGCGTA 749  
 Db 750 TATTCGTAGAGCTTTGGCGGATATGGCAATGATGACGAGAAACCGTGGCTCTACG 809  
 Db 810 CAATTCGCGAGAGTTCCTGGCGCATGGCCATGAACGACGTGGAACAGCGCGGCTGATCGT 869  
 Qy 810 GCGTGGCACACACTCGGAAAAACCCATATGCTGCGCATGGGAGAAATATGTGGCGCG 869  
 Db 870 CGGCGGTGCATCTTTTGTGTGAGACCCATGAGCGC---CGGCGCGCGCGCTGTGGTGGCGCC 926  
 Qy 870 AGACCTTCGCGCGCGAGGTATTCAGAAATAGAGCTGGGTGGGAGAAACACTACGCGAC 929  
 Db 927 CGAAACCGGAGCGTCTCCCTCGGAGCAGATGGCGTTGGCGTGGAGAGCTCGTATGCGAC 986  
 Qy 930 CGGACACCGTGGGATACCATCCAGTGGACTAGAGAGGCGCTGGACCCAGACCCCTAC 989  
 Db 987 CGGACCGGTAGAGACCGGCTACACCCCGCATCGAGGTGTATGGAGGACACCCCGAC 1046



QY	990	TCAGTGGAGCAATACCTTTTGAACACTCTTTCTGCTACAGCTGGCGAGCTTACCAAG	1049	Db	2121	TGGCAAGGTGAAGTGGACGGCAGCGGCTGGACCTTGGCTTTCGGCTCCACCTCGGAGTT	2186
Db	1047	GAATATGGGACACAGTTCCTCTCGAGATCTGCTACGGCTACAGATGGGAGCTGGACGAAG	1106	QY	2127	AGAGAGCTTCCGACAGAAATGTACGGCTGTGACAGATCTGAGAAAAGATTTGTATAAGATTT	2186
QY	1050	TCAGCTTGGAGCTTATCATGTGGAAACCAAAAGACGGTGCGGGGCTGGCACATATCCGGA	1109	Db	2181	GGCGGCTTTTCGAGGCTATAGCGCGGATGAGCGCAGCGAAGTTCGTGCGAGGATTT	2240
QY	1107	CCCTCGTGGCTGTGGCAATGACACCCGACAGCGCGCGGCTGGCGACCATATCCCGGA	1166	QY	2187	TGTGAGAGCTTGGGCGCAAAAGTATGAGCTTGGACCGGGTGTGATCTG	2232
Db	1110	TGCGCATGATGATGCGACAGCTCCCATTTATCTCACTACGAGCTGGCGCTGCG	1169	Db	2241	CGTGGCTGCTGGGACAGGTGATGAACCTCGCAAGTTTGACGTG	2286
QY	1167	CCGCTTGG---GCGGGCCAGGGCGCTCCCGGACGATGCTGACCCTGAACCTCTCGCTGCG	1223	RESULT 9			
QY	1170	CATGACCCCTGATACGAAAAAATTTCTCGACGGTACTATGAAAACCTGATGATGTTGC	1229	ANT89967			
Db	1224	GGTGGATCGATCATATGAGCGGTACAGCGTCTGCTGGCTGGACACCGAGAAATGGC	1283	XX	ANT89967		
QY	1230	AGATGCTTTTCGGAAGCATGGTACAACTGACACACAGATGATGGAGCAAAAGGTGCG	1289	DT	20-MAR-1998	(first entry)	
Db	1284	CSACGAGTTGCGCAGCGCTGTACAAAGCTGATCCACGAGCATGGGTCCGCTGGCAG	1343	XX	Mycobacterium tuberculosis partial katG gene.		
QY	1390	TACTCTGGGACCAAGATGGCTTCAGGAGACCTCATCTGGCAGACCTGATACAGATGT	1349	XX	Tuberculosis; katG gene; Isoniazid resistance; INH;		
Db	1434	ATACCTTGGGGCGGTGGCCCAAGCAATGATGAGCGCTGCTGGCAGGATGCTTCGGGT	1403	XX	Isoniazidic acid hydrazide; IS.		
QY	1350	AAAGCCATCTCTGTGTAGAACCAAGATATGAGCGCTTAAGAGCCAAATCTCGGATC	1409	OS	Mycobacterium tuberculosis.		
Db	1404	CAGCCACGAGACTCGTGGCGAGGCGGAGATTGCCAGCTTAAGAGCCAGATCCGGGCATC	1463	XX	Key Location/Qualifiers		
QY	1410	GGCATGACGATGAGCGATGGTATAGCACGGATATGCGGTCTGTGACTACTTTTAGAAA	1469	FT	CDS	85..2235	
Db	1464	GGATGTCTCTCAAGCTATTTTCGAGGATTTTCGAGGATGCGGCGCTGCTGCTGGTGG	1523	FT	/*tag= a	/*tag= a	/*tag= a
QY	1470	CTCTGACAGCGGGGGTGGCCACGCTGACATATACGATGATGCTGGCCGACAAAGACG	1529	FT	/product=	/*tag= a	/*tag= a
Db	1524	TAGCCAGACAGCGCGCTGGCGAGCGGCTGTCATCGCTGCGCTGAGCGACAAAGTGGG	1583	PN	/note=	partial cds	partial cds
QY	1530	GGAGCTTAAACAGAGCTCA---GCACTTGGCGAGGTACTCAAAACATAGAGGTATCCA	1586	PD	US5688639-A.		
Db	1584	GGAGTCAACGACCCGACGGGATCTGCGAAGCTCTGCGAGCTCTTCCAGCTTGGGAGATCA	1643	XX	18-NOV-1997.		
QY	1587	GGAGGACTTTACAGCGGCAATCAGATACAAAGCAGTATGCTGGCGACCTGATTTG	1646	XX	18-APR-1994; 94US-0228662.		
Db	1644	GGAGTCAATCAACTCGCGGCGCGCGGAGACATCAAAAGTGTCTTCCGCGACCTCGCTGT	1703	XX	18-APR-1994; 94US-0228662.		
QY	1647	CTGCGCGCGGGGGTGTAGAAAAGCTTGGAAAAGTCTGGCATGAGTGGTGCAGGT	1706	PI	(MAYO-) MAYO FOUNDATION.		
Db	1704	CTCTGGTGGCTGTGGCCATATAGAGAACGCAAGCGGCTTGGCCACACATCAACGCT	1763	XX	Cockerill FR, Kline BC, Uhl JR;		
QY	1707	GCTTTCAACCCGGGACAGCGGATGSCACCGCTGACCAACCGATGTGGAGCTTTGGA	1766	XX	WPT: 1998-007975/01.		
Db	1764	GCTCTTACCGCGGCGCGCGAGTGGCTGGCGAGCAACACGAGTGGATCTTTGTC	1823	XX	Determination of isoniazid sensitivity of Mycobacterium tuberculosis strains - by restriction length polymorphism analysis of katG gene		
QY	1767	AGCATATAGACCGGCTGCTTGAAGATACATTAACGAGCAATMAAGTATC	1826	XX	Claim 1: Column 9-12; 18pp: English.		
Db	1824	CGTGTGAGCCCAAGCCAGATGGCTTCGAAACTACCTCGGAAAGGCGACCCGTTGCG	1883	CC	This partial DNA sequence encodes the katG gene of Mycobacterium		
QY	1827	CGCTGAGGAATGCTGTAGACCGGGCGAGCTTCTGCTGGTTCGCGACAGAAATGAC	1886	CC	identify strains of M. tuberculosis (see abstract) used in a novel method to rapidly		
Db	1884	CGCGGCTACATGCTCTGCAAGGCGACCTGCTAGCTCATGTCGCTCGATGATGAC	1943	CC	(INH, Isoniazidic acid hydrazide). The method involves the use of		
QY	1887	TGCTTTGGTAGCGGATATGGTGTACTTGGGCAACATACAGCGTTCGCGATGGAT	1946	CC	restriction fragment length polymorphism (RFLP) analysis to determine		
Db	1944	GGTGTGGTAGTGGCTTGGCGGGTCTCTGGCGCAACTACAGCGCTTACGCTGGCGGT	2003	CC	if a NotI-MspI restriction site is absent in the DNA of the strain at		
QY	1947	GTTTACAAATAGCGGGTCAAGCTATTCATGATCTTTGTAACCTGCTAGACATCA	2006	CC	the codon corresponding to codon 463 of a M. tuberculosis katG gene		
Db	2004	GTTACAGAGGCTCCGACTGACTGACACAGCACTTCTTGGAGCTCTTCGATGAGG	2063	CC	consensus sequence (see AW31343). The absence of the site indicates an		
QY	2007	CATTAATGSGAGCGATGATACAGACAGATTTTTCGAGCGAGAGCTTCAAC	2066	QY	30	TAGCTTGAACCAACTCGCGGAAATGCCCTTTAGCGAGGTTCGCTTACCAAG	89
Db	2064	TATCACTGGGAGCCCTGGCCAGCAGATGACGGGACCTACACGGGCAAGATGCG---AG	2120	Db	33	TACAGAACCAACCAACGAGCGCTGATGACAGCGCTGCTCGCTGCTCATATGAA	92
QY	2067	TGGCAGATTAAGTGGAGTGGCACCGCGGTATGACCTGATCTTGGGATCCATTCGAGCT	2126	QY	90	TGCGAGTGGCGGCGCAAAAAACAGGAGTGTGGTGGCCCAACATCTCTCAACCTCGGCACTT	149

Db 93 CCCGCTGAGGGGGGGAACAGAGCTGGTGGCCMACCGGCTCAATCTGAGGTACT 152  
 QY 150 ACCECAACATCATCCCTATCCAGACCAACAGACCGGATTTGCTATCTGCGAAGGTT 209  
 Db 153 GCACCAAAACCGCCCGCTGCTGACCGCATGGTCCGGCGCTTCGCTATATCGCGGAGGT 212  
 QY 210 TAAGAAGCTAGATCTGGCAGCCGGTTTAAAGAGGACCTGGCAGCGCTAATGACAGATATACA 269  
 Db 213 CGCGACATCTGACCTGTGACGGCTCTGACGGGGGACATCAGGAGAGTGTATACCACTGGGA 272  
 QY 270 CTTGCGCTGGTGGCAGATATGCGTGAATATGGCCCTCTCTTATACCGATCGGTGGGCA 329  
 Db 273 GCGGTGGTGGCCCGCAGCTAGCGGCACCTAGCGGGCCGCTGTATTACCGGATGGCGTGGCA 332  
 QY 330 CAGCGCCGGCAGCTACCGGTATCGGTGATGGCCGCTGGTGGCGGTGCTCCGGCTACAGCG 389  
 Db 333 GCTCTCGCGGACCTACCGCATCTACAGCCGCGCGCGCGCGCGGGCGGCGATCGCGG 392  
 QY 390 CTTTGGCCGCTCTCANTAGCTGGCCAGACATGCCAATCTGGATNAGGACGGTCTCTCT 449  
 Db 393 GTTTCGCGCGCTTANCAAGCTGGCCGACACAGCCAGCTTGGACNAGGGCGCCGCTCTCT 452  
 QY 450 TTGGGCCCATCMAACAAATACGGTGCMAAATCTCTCTGGCGGCTCTAATGATATCAC 509  
 Db 453 CTGCGGTGACAGAACTATGCGCAAGAGCTTCATGGCGGACCTGATGTTTTTGGC 512  
 QY 510 AGGAAGCTAGCTCTGGAATCTTGGCTTTAAACCTTTTGGTTTTCGAGTGGCGAGCG 569  
 Db 513 CGGCACCTCGCGCTGGAATCGATGGGCTTCAAGACGTTGCGGTTGCGCTTCGCGCGGT 572  
 QY 570 AGATGATGGAGCTGMAAGAGATGATGATGGGAGCAGAAACCGATGGCTGGGAGA 639  
 Db 573 CGACCACTGGGAGCC---CGATGAGTCTATTGGGCGAAGNAGCCACTTGGCTGGGCG 639  
 QY 630 CAAGCGCTATGAGCTACCGCAGAGCTTCGAAATCTCCCTGGAGCGCTACAAATGGACT 689  
 Db 630 TGAGGCTTACACGGCTATCGGAGGATCTGGGAACCCCTGGCCGCTGCAGATGGGCT 689  
 QY 690 GATCTAGTGMAACCGGAGGGCGAGCGCACCCGACCCACCGCCGCGCGGTGCA 749  
 Db 750 TATTCTGTGAGCTTTGGCCGGAATGGCAATGATGAGGAAGAACCTGGCTCTCATAGC 809  
 QY 750 CATCTGGAGAGTTTTCGGCGCATGGCCATGACAGCGCTCGCAACAGCGCGCTGATGT 809  
 Db 810 GGTGTGCAACACTCTGGAACAAACCTCGTCTCGCATGGCGGAGAGAAATATGCGGCG 869  
 QY 810 CGCGCGTCACTTTTGGTTAGAACCCATGCGGC---CGGCGCCGCGGATCTGCTGGCGCC 866  
 QY 870 AGAGCTGCGCCCGCAGGTATTGAAGAAATGAGCTGGGCTGGGAAMACACCTACGGCAC 929  
 Db 867 GMAACCTAGGCTTCTCCGCTGGAGCAGATGGGCTTGGGCTGGAGAGCTCTGATGGCAC 926  
 QY 930 CGGACAGGTGCGCATACATCAACAGTGCATTAAGAGCGCTGCGACCAAGACGCTAC 989  
 Db 927 CGGAACCGGTAGAGAGCGGATCACACGCGCATCGAGTGGTATGAGCAACACCCGNC 986  
 QY 990 TCAATGAGCATATCTTTTTGAAACCTCTTTGGTTACGATGGGAGCTTACCAAG 1049  
 Db 987 GAATGGGACACAGTTTCTCTCGAGATCTGTACGGCTACAGTGGGAGCTGACGAG 1046  
 QY 1050 TCCAGCTGGAGCTATGTGGAAGCAAGNAGAGCTGGCGGCTGGCGCATACGGA 1109  
 Db 1047 CCTCTGGCGTTTGGCATACACCGCAAGACGCGCGCGTGGCGCGCCACATCCGGA 1106  
 QY 1110 TGCACATGATCCCAAGATGCGACGCTCCATTTATGCTCACTACGGAACCTGGCGGTGG 1169  
 Db 1107 CCGGTTGG---GCGGCGAGGGCGTCCCGGAGGATGCTGGCAGCTCTGCTGGG 1163  
 QY 1170 CATGCGCCTGNTACGAAAAATTTCTGCAGGCTACTATGAAACCCCTCATGAGTTTGGC 1229

RESULT 10  
 AAT89968

Db 1164 GGTGGATCCGATCTATAGCGGATACGGCTGCTGCTGSAACACCCGAGMAATTGGC 1223  
 QY 1230 AATGCTTTTCGGGAAGCATGGTACAACTGACACAGATATGGGCAAAAGGTGGC 1289  
 Db 1224 CGACGAGTTTCCCAAGGCTTGTGACGCTGTCCACGAGACATGGCTCCGTTTGGAG 1329  
 QY 1290 CTACCTGGGACAGAGAGCTCTAGAGAGCACTCATCTGCGAGCAGCTATACAGATGT 1349  
 Db 1284 ATACCTTGGGCGCTGGTCCCCCAAGCAGACCTGCTGTGCGAGGATCGGCTCCGTGGGT 1343  
 QY 1350 AAGCACTCTTGTGTAGAGAAAGAGATATGAAAGGGCTTAAAGCCMAATCTGGGATTC 1409  
 Db 1344 CAEGCAGACCTCTGCTGGCGAGCTTGGCAGCTTATAGACCAAGATCOGCGCATC 1403  
 QY 1410 GGAAGCTAGACGTAAAGAGAGTGTGTAACAGCGCATGGGCTCTCATCTACTTTAGAAA 1469  
 Db 1404 GGAATGATGCTCTCAAGCTAGTTTTCGACCGATGGCGGGCGGCTGCTTCCTGG 1463  
 QY 1470 CTGTGACAGAGGCTGGCGGCTGCGACGTCGATGATGAGCTGGCGGCTCAAAAAGAGTG 1529  
 Db 1464 TAGCGACAGCGCGGCGGCCACGCTGCTGCTGCGCTGCGCCACCAAGTGGGATG 1573  
 QY 1530 GGAAGTAAACACCTCCA---GCACACTTCCAGGGGTACTCAAACTAGAGAGTATCCA 1586  
 Db 1524 GGAAGTACAGACCCCGCAGGGGATGTTCGCGCAAGGCTCATTCGCACTGGAGAGATCCA 1583  
 QY 1587 GGAAGCTTTTACCAGCGGCAATCAGATATGAAAGCAGTATGCTTGGCGGCGATGTTGT 1646  
 Db 1584 GGAATCATCACTCCGCGCGCGGGAACATCAAGTGTCTTTCGCCGCTGCTGCT 1643  
 QY 1647 GCTGGCGGCTGCTCGGGTGTAGAAAAGCTGCAAAAGATGCTGGCCATGAGGTGCAAGT 1706  
 Db 1644 GCTGTTGGCTGTGCGGCAATAGAGAAAGCAGCAGCGGCTGGCCACACATCAGGCT 1703  
 QY 1707 GCTCTTCAACCGGAGCGGAGTGCACCGCTGAGCAACCGATGCGAGAGTTCGA 1766  
 Db 1704 GCTCTTCAACCGGCGCGCACGATGCGTGCAGGAACACCGAGTGGATCTTTTGC 1763  
 QY 1767 AGCACTAGAGCCAGCGCTGACGGCTTTAGAAACTACATTAACCGGAGCATTAAGTATC 1826  
 Db 1764 GCTGCTGAGGCCAAGCAGTGGCTTCCGAATCTCTCGAAGGGCACCGGTTGGC 1823  
 QY 1827 GCGTGAAGAAATGCTGTAGACCGCGCGCACCTTCTGCTGCTTTTCGGCAGCAAGATGAC 1886  
 Db 1824 GSCCGAGTACCTGCTCTGACAGCGGAACCTGCTACGCTCAGTGGCCCTGAGATGAC 1883  
 QY 1887 TGTCTGTAGCGGTATGCTGTATGCGGCACCACTAGACGCGTTCGACGATGGAGT 1946  
 Db 1884 GCTCTGTGTAGTGTGCTGCTGCTGCG---GCANCTACAGCGCTTACCGTGGGCGT 1942  
 QY 1947 GTTTTAAATAAGCCGGGTGAGTATCAATCAATGACTTCTTTGTAACCTGCTAGCTCAAA 2002  
 Db 1943 GTTTCACGCGGCTCCGAGTCTGACCAACGACTTCTCTGATGACTGCTGCACATGGG 2002  
 QY 2007 CACTAAATGCGGACCCAGCATGATACAGCAAGTTTTTGAAGCGAGACTTCAAAAC 2066  
 Db 2003 TTTCACTTGGAGGCTCTCCGACAGATGAGGGGACCTACGAGGGCAGGTGGC---AG 2059  
 QY 2067 TGGCAGAGTAAAGTGGAGTGGCAGCCCGGTGAGCTGATCTTGGATCAATTTCCGAGCT 2126  
 Db 2060 TGGCAAGGTGAATGTGAGCCGCGCCGCTGGACCTGCTTTCGCGTCCAACCTCGAGTT 2119  
 QY 2127 AAGAGCCTCGCAAGTGTAGCGCTGTGCGAGATCTTGAAGAAAGTTTGTGAAGATTT 2186  
 Db 2120 GCGGGCGCTTGTGCGGTCTATGCGGCGCATGACGCGAGGCTGCTGCGAGACTT 2179  
 QY 2187 TGTGAAGGCTCGGGCAAGATGAGCACTGGACCGGCTTTTATCTG 2233  
 Db 2180 GCTGCTGCTCGGACAAAGGTGATGAACCTCGACAGGTTGACGCTG 2225



Key	CD5	Location/Qualifiers	1979-1986	1987-1993	1994-2003	2004-2011	2012-2019	2020-2027	2028-2035	2036-2043	2044-2051	2052-2059	2060-2067	2068-2075	2076-2083	2084-2091	2092-2099	2100-2107	2108-2115	2116-2123	2124-2131	2132-2139	2140-2147	2148-2155	2156-2163	2164-2171	2172-2179	2180-2187	2188-2195	2196-2203	2204-2211	2212-2219	2220-2227	2228-2235	2236-2243	2244-2251	2252-2259	2260-2267	2268-2275	2276-2283	2284-2291	2292-2299	2300-2307	2308-2315	2316-2323	2324-2331	2332-2339	2340-2347	2348-2355	2356-2363	2364-2371	2372-2379	2380-2387	2388-2395	2396-2403	2404-2411	2412-2419	2420-2427	2428-2435	2436-2443	2444-2451	2452-2459	2460-2467	2468-2475	2476-2483	2484-2491	2492-2499	2500-2507	2508-2515	2516-2523	2524-2531	2532-2539	2540-2547	2548-2555	2556-2563	2564-2571	2572-2579	2580-2587	2588-2595	2596-2603	2604-2611	2612-2619	2620-2627	2628-2635	2636-2643	2644-2651	2652-2659	2660-2667	2668-2675	2676-2683	2684-2691	2692-2699	2700-2707	2708-2715	2716-2723	2724-2731	2732-2739	2740-2747	2748-2755	2756-2763	2764-2771	2772-2779	2780-2787	2788-2795	2796-2803	2804-2811	2812-2819	2820-2827	2828-2835	2836-2843	2844-2851	2852-2859	2860-2867	2868-2875	2876-2883	2884-2891	2892-2899	2900-2907	2908-2915	2916-2923	2924-2931	2932-2939	2940-2947	2948-2955	2956-2963	2964-2971	2972-2979	2980-2987	2988-2995	2996-3003	3004-3011	3012-3019	3020-3027	3028-3035	3036-3043	3044-3051	3052-3059	3060-3067	3068-3075	3076-3083	3084-3091	3092-3099	3100-3107	3108-3115	3116-3123	3124-3131	3132-3139	3140-3147	3148-3155	3156-3163	3164-3171	3172-3179	3180-3187	3188-3195	3196-3203	3204-3211	3212-3219	3220-3227	3228-3235	3236-3243	3244-3251	3252-3259	3260-3267	3268-3275	3276-3283	3284-3291	3292-3299	3300-3307	3308-3315	3316-3323	3324-3331	3332-3339	3340-3347	3348-3355	3356-3363	3364-3371	3372-3379	3380-3387	3388-3395	3396-3403	3404-3411	3412-3419	3420-3427	3428-3435	3436-3443	3444-3451	3452-3459	3460-3467	3468-3475	3476-3483	3484-3491	3492-3499	3500-3507	3508-3515	3516-3523	3524-3531	3532-3539	3540-3547	3548-3555	3556-3563	3564-3571	3572-3579	3580-3587	3588-3595	3596-3603	3604-3611	3612-3619	3620-3627	3628-3635	3636-3643	3644-3651	3652-3659	3660-3667	3668-3675	3676-3683	3684-3691	3692-3699	3700-3707	3708-3715	3716-3723	3724-3731	3732-3739	3740-3747	3748-3755	3756-3763	3764-3771	3772-3779	3780-3787	3788-3795	3796-3803	3804-3811	3812-3819	3820-3827	3828-3835	3836-3843	3844-3851	3852-3859	3860-3867	3868-3875	3876-3883	3884-3891	3892-3899	3900-3907	3908-3915	3916-3923	3924-3931	3932-3939	3940-3947	3948-3955	3956-3963	3964-3971	3972-3979	3980-3987	3988-3995	3996-4003	4004-4011	4012-4019	4020-4027	4028-4035	4036-4043	4044-4051	4052-4059	4060-4067	4068-4075	4076-4083	4084-4091	4092-4099	4100-4107	4108-4115	4116-4123	4124-4131	4132-4139	4140-4147	4148-4155	4156-4163	4164-4171	4172-4179	4180-4187	4188-4195	4196-4203	4204-4211	4212-4219	4220-4227	4228-4235	4236-4243	4244-4251	4252-4259	4260-4267	4268-4275	4276-4283	4284-4291	4292-4299	
-----	-----	---------------------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	--

Db 2302 CGCTCGCGGCACTTACCGCATCCAGACCGCCCGCGCGCGCGCGCGCATGTCACGCG 2361  
 QY 390 CTTTGGCGCTCTCAATAGCTGGCGAGCAATGCCAATCTGGATAMAGCAGCTTGCTCT 449  
 Db 2362 GTTGGCGCGCTTTACACAGCTGGCGGCAACAGCCAGCTTGGACAGCGCGCGCGCTCT 2421  
 QY 450 TTGGGCCCAACAAATATGCGTGGAAATCTCTCTGGGGGATCTTAATGATCTAC 509  
 Db 2422 GTGGCGGCTGANGAGATAGCGGCAAGCTCATATGGCGGACCTGATGTTTGGT 2481  
 QY 510 AGGAAC--GTAGCTCTGAAACTTGGGCTTTAAACTTTGGTTTCAGGTGGCGACA 567  
 Db 2482 GGGCAACCGCTCGCAATGATGGGCTTCAAGACGTTCGGCTTGGGCT 2541  
 QY 568 GCAGATGATGGGCGCTGAGAGAGATATCTGGGGAGCAGNACCGAATGCTGGGA 627  
 Db 2542 GCACAGTGGAGACC----GATGAGTCTATTGGGGCAAGGACCACTGGCTGGCG 2596  
 QY 628 GACACGGCTCTGAAAGCTGACGAGAGCTCGAAATCCCTCTGGAGCGGCTACAAATGGGA 687  
 Db 697 GATGA--CGGTTCAGAGCTGATGGAGTCTGGAGACCGCTGCGCGGTGGCAGATGGG 2653  
 QY 688 CTGACCTGTAAACCGCAAGCGAGCTATCTGGAGCGAGAGCTATGCTCTGTGGCGT 747  
 Db 2654 CTGACTAGCTGACCCGAGGCGCGCAACGCAACCGGACCCCATGCGCGCGCGCTCT 2713  
 QY 748 GATATCTGAGAGCTTTGGCGCGAATGGCAATGAATGACGAGAAACCGTGGCTCTCAT 807  
 Db 2714 GACATTTGCGAGGCTTGGCGCGATGGCCATATGACGAGTGCAGAACGGCGCGCTATC 2773  
 QY 808 GCGGTGGACACCTCTGGAAACCGATCTGCTCGGATGCGGAGCGAGAAATATGGGCG 867  
 Db 2774 GTGCGCGTCACTTTCGTTAGAACCATGG--CGCGCGCGCGCGGATCTGTGGCG 2830  
 QY 868 CGAGAGCTCGCGCGCAGGTATTGAAGAAATGAGCTGGGCTGGGTGGGAAACACCTACGCG 927  
 Db 2831 CCGGAACCGAGGCTGCTCGCTGGAGAGATGGGCTTGGGCTGGAGAGCTGTATGGC 2890  
 QY 928 ACCGACAGCGGTGGGATCACTACCTAGTGAAGCGGCTCGGACGACGCGCT 987  
 Db 2891 ACCGACAGCGGTAAAGAGCGGATACCCAGCGCATCGAGTCTGTATGGACGAACCCCG 2950  
 QY 988 ACTCAATGGAGCAATACCTTTTGAACACCTCTTGGTATCAGAGTGGGAGCTTACAAA 1047  
 Db 2951 ACGAATGGGACAGATTTCTTCGATGCTCTAGGCTACGAGTGGGAGCTGACGAG 3010  
 QY 1048 GTGCACTGGAGCTTATCAGTGGAAACCAAGAGCGTGGCGGCTGGCACTACCG 1107  
 Db 3011 AGCCTCTGGGCGTTGGCAATACACCGCAAGGACGCGCGCGTGGCGGCACTACCG 3070  
 QY 1108 GATGCACATATCCCAAGCAGTGGAGGCTCCATTTATGCTACTACGAGACCTGGGCTG 1167  
 Db 3071 GACGCTTGG--CGGCGAGGCGTCCCGATGATGCTGGCAGTACTGCTCTGCTG 1227  
 QY 1168 GCATGAGCTGATATGAGAAATTTCTCGAGCGCTACTATGAAACCTGTAGTGT 1277  
 Db 3128 CGGGTGGATCGATCTATGAGCGGATCAGCGCTGCTGGCTGGACACCCCGAGATTTG 3187  
 QY 1278 CGAGATCTTTGGGAAAGCATGGTACAACTGACACACAGAGATATGGGACCAAGGTTG 1287  
 Db 3188 GCGAGAGTTCGCAAGGCTGGTACAGCTGTATCCAGAGACATGGGCTCGGTGG 3247  
 QY 1288 GCTACTCTGGACACAGAGTGGCTCAGGAGACCTCATCTGGCAGACCTATACAGAT 1347  
 Db 3248 AGATACCTTGGCGGCTGGTCCCAACGACAGACCTGCTTGGCAGGATCGGTCCTCGG 3307  
 QY 1348 GTAAAGCATCTCTTGTAGACAAACGATATTTGAAGGCTTAAGGCCAAATCTCGAA 1407  
 Db 3308 GTGAGC--ACAGCTCTGGCGAGCGAGATTCGACGCTTAAAGAGCCAGATCGGGCA 3364  
 QY 1408 TGGGAGCTTGTAAAGCGCTGTAAAGACGCGATGGGCTCTCTGCATCTACTTTAGA 1467

Db 3365 TCGGGATGTAGTGTCTACAGCTAGTTTGCAGCGCATGGCGGGCGGCTGTGTCTCCGT 3424  
 QY 1468 AACTCTGACACGCGCGCGCTGCGCAAGCTGTGACGCTTACGACTTGGCCGCAACAAAGAC 1527  
 Db 3425 GGTAGCAGACGCGCGCGCGCAACGCTGGTGCATCTCCGCTCGACACCAAGTCGGG 3484  
 QY 1528 TGGGAAGTAAACAACCTCTAGCAACTTGGCAGGGTACTCAAAACACTAGAAGTATCAG 1587  
 Db 3485 TGGGAGGTCAACGACCCGCGAGATCTTGGCGACGCTATTTTGCACCTCTGAAGATATCAG 3544  
 QY 1588 GAGGACTTTTACGAGCGCAATCAGATACAAAGCAGTATCTTGTGGCGAGCTGATGTTG 1647  
 Db 3545 GAGTCAATCA-----CTGCGCGCGGGAACATCAAGATGTCTTTCGCGAGACTGTCTG 3598  
 QY 1648 CTGGCGGCTGTGGGGGTGTAGAAAAAGTGTCAAAAAGTCTGGCATCTAGGTGTCAGGTG 1707  
 Db 3599 GTGCTGGTGTGTGGCCATCTAGAAAGACAGAAAGCGCTGTGCTCAACATCTTCCAG 3658  
 QY 1708 CTTTCAACCGCGGAGCGGATGCCACCTGAGCAAAACGATGTGGAAGCTTTCGAA 1767  
 Db 3659 CCCTTCAACCGCGCGCGCGACAGTCTGCTGGCAGCAACAAACGACGTGGAATCTTTGCC 3718  
 QY 1768 GCATAGAGCAGCGGCTGACGGCTTTAGAACTTATTAACCGGAGCATTAAGTATCC 1827  
 Db 3719 GTCTGGGCGCCCAAGAGATGTCTCCGAACCTCTCGAAAGGACACCTTGTGGCGG 3778  
 QY 1828 GCTGAGGAAATCTGCTGAGCGGGCGAGCTTCTGCTGTTTGGCAGCAAGATCACT 1887  
 Db 3779 CCGAGTACATCTGCTGTGACAGGCGAAGCTCTTACGCTCAGTGCCTCTGATGAGG 3838  
 QY 1888 GCTTTGGTAGCGGTATGCGGTCTACTGGCACCACTACGACGCTTGGCAGCATGAGGTG 1947  
 Db 3839 GTCTGGTGTGTGTGCTGTGGGGTCTGCGCGCAACTACAGCTTACGCTTGGCGGTG 3898  
 QY 1948 TTACAAATAAGCGGGTACGCTATTCATGATCTTGTGTAACCTCTAGAACCTCAAC 2007  
 Db 3899 TTACCGAGGCTCTCCGAGCTACTGACCAACGACTCTTCTGTAACCTCTGACATGGT 3958  
 QY 2008 ACTAATGGCGAGCAGCGATGATGATACAGAAAGTTTTCAGCGAGAGACTTCAAACT 2067  
 Db 3959 ATCACTTGGAGGCTTCCGCGCAGATGATGAGGAGCTTACGAGGAGATGGC---AGT 4015  
 QY 2068 GCGCAAGTAAAGTGGAGTGGCCAGCCGGGTAGAGCTGATCTTCGGATCCAAATTCGAGCTA 2127  
 Db 4016 GCAAGGTGAAGTGGACGCGCGGTGAGCTTGTCTTGGGCTCACTCGAGATTTG 4075  
 QY 2127 GAGAGCTTCCGAGATGATGAGGCTGTGCGATCTGAGCAAGTCTGAGAAAGTGTTAAGATTT 2187  
 Db 4076 CGGCGCTTCTGAGGTCTATCTGCCCGATGACGCGCAGCGCAAGTGTGTGACAGATTC 4135  
 QY 2188 GTGAGGCTGGGCAAGATATGAGCTGGAGCGGTTTGTATCTG 2232  
 Db 4136 GTGCTGTGGGGAAGGTATGAACCTCGACAGGTTTCGAGGTG 4180

## RESULT 12

AAI16969

 ID AAI16969 standard; DNA; 4795 bp.  
 AC AAI16969;

XX AAI16969;

XX 20-MAR-2003 (updated)

DT 11-MAY-1995 (first entry)

DE M.tuberculosis katG gene.

XX katG; catalase; peroxidase; enzyme; isoniicotinic acid hydrazide;

XX isoniazid; INH; diagnosis; resistance; ss.

XX Mycobacterium tuberculosis.

XX Key

XX Location/Qualifiers

FT 1979..4186









Best Local Similarity 59.5%; Pred. No. 1.8e-199; Matches 127; Conservative										
QY	99	CGGCACCAAAACAGGATTTGGTGGCCCAACATGCTCAACCTCGCATCTTACGCGCACAA	158	DB	1164	GCTGGCATCGACCGCTGATTTACGAAAAATTTCTCGACGGTACTATGAACAACTCTGATGA	1223			
DB	150	CGGTACTTCCACAAAGTTTTGGCGGAAGGGTTGAACCTGGATTTTGTCTATCAGCA	209	DB	1227	GATAAAGTAAATCGACCTATCGCGCTATCTGCGAAAAATCTATGCGCGCATCTCTGAGTA	1286			
QY	159	TTCTATCGCTATTCGGACCAACAGCCGGATTTTGACTATGCGGAAGCTTTAAGAGCT	218	QY	1284	GTITTCAGATGCTTTTGGCAAAAGATGGTACAACTGACACAGAGATATGGGACAAA	1283			
DB	210	AGATGCGCAATATCAAGACCCGATGATTCGGATTTCACTACTCGGTGAGAGATAGCGNAGCT	269	DB	1287	CTTCACAAACCTTTTGGGAGAAAGGGTGTCTAAGCTGACGACCGCTGACCTGGGCGGAA	1346			
QY	219	AGATCTGGCAGCGTAAAGAGACCTGGCAGCGCTAATGACAGATTCACAGAGCTGGTG	278	QY	1284	GGTGGCTTACCTGGCAGAGAGTGGCTCAGGAGAACTCTATCTGGCAAGACCTCTATACC	1343			
DB	270	CGATTTTCGACCGCGCTGAAGAAAGATGTCCACGGCTGTATGACCGTACGCAAGATGGTG	329	DB	1347	ATCACGTACATCGCCCGGAAGTGGCGGACAGAACTGATTTGGCAAGCCGATTC	1406			
QY	279	GGCAGAGATACAGGTCTATATGCGCCCTCTTTTATACGATAGGGTGGGACAGCGCG	338	QY	1344	AGATGTAAAGCATCTCTCTGTGTAGAGAAAMAGCTATTTAGAGCTTAAAGCCCAAAATCCT	1403			
DB	330	GGCGCTGACTGGGCGCACTACGCGGTTTGATGATCCGTATGGCTTTGGGCACTCGCGTGG	389	DB	1407	GGCAGGTGAC-----ACCGACTATGCGGAAGAAGTGGTCAAGCAAGAAATATGC	1454			
QY	339	CACCTACCGTATCGGTGATGCGCGTGGTGGCGGTGGCTCGGCTCACACGCGCTTGGCGC	398	QY	1404	GGATCGGCACTTGAAGGTAAAGCAGCTGTGTAAGCAGCGGTCTTGCATCTACTTT	1454			
DB	390	CACCTACCGTATGCTGATGGCGGTGGGCGGGTGGTACGGAAAGCAGCGCTTTCGACC	449	DB	1455	ACAAAGTGGCGTGAAGTATGAGATGGTCTCACCGCTTGGGACAGTGGCGCTACTTA	1514			
QY	399	TCCTATAGCTGGCCAGACATGTCATCTGATTAAGCAGCTTGCTCTCTTTGGCCCAT	458	QY	1464	TAGAAATCTGACAGCGCGCTGCCACAGCTGTGACAGCTATACAGCTGGCGCCCAAAA	1523			
DB	450	GCTCATCTCTCGCGGCAACAGCTCAGCTGTGATTAAGCGCGCTCTGCTGTGGCGAT	509	DB	1515	TCGGGTTCGATATGCGGGGGTGGTCAAGCGTCCCGCATCTGCTTGGCGCCACAGAA	1574			
QY	459	CAACCAAAATACGCTCGAAATATCTCTGGCGGGATCTATGATGATCTACAGGAAGCT	518	QY	1524	AGACTGGGATTAACAAACCTCTCAGCAACTTGGCAGGCTACTCAAAACATAGAGATAT	1583			
DB	510	CAGAGAAATACGGCAGCAACAAATATCAGCTGGGCGAGACTGATGATCTGGCTGGCACCT	569	DB	1575	CGAGTGGCAGGCGACAGCGCGCGCTGGCGAAGTGC-----	1616			
QY	519	AGCTCTGGAATCTATGGGCTTAAACATTTGGTTTGGCAGGTGGCAGAGCAGATATG	578	QY	1584	CAAGAGGAGCTTTTACACAGGCGCAATCAGATACAAAGACGATATCGTTGGCGACCTGAT	1643			
DB	570	GGCTTATGATGCTTACCTGCTTACGGCTCTCTTCTGGCGGTGCTGATTTTG	629	DB	1644	TTGTGCTGGCGGCTTGGCGGTGTAGAAAGCTGCAAAAGATGCTGGCATGAGTGCA	1703			
QY	579	GGAGCTCGAAGAGATGTATCTGGGAGCAGAAACCGAATGGCTGGGA-----GACAA	632	DB	1671	CGTCTTGGCGGCTGAGCTTAGGCTTCGAAAGCCGCGAAGCAGAGTACGATGCG	1730			
DB	630	GGACCGGAAAGATATCTATCTGGGCTGACGAAAGAGTGGCTGGCACCTCTTCAGCA	689	QY	1704	GGTGGCTTTCAACCGGGAGGAGGCGACCGCTGAGCAACCGATGTGGAGCTTT	1763			
QY	633	GGGCTGATGAGGTGACCGAGAG-----TCGAAATGCCCTGGGAGCGCTACAT	683	DB	1731	CGTCTCTCTCTGAAAGCGCTGGCGATGCGACCGCGAGATGACGCGAGACTCTTT	1790			
DB	690	ACGCTACGGCGAGCTGAACACCGCAGAGACCATCGAANAACCGCTGGCGGTGTCCAAAT	749	QY	1764	CGAGCACTAGACCAAGCGCTGACGCTTTAGAACTACATTAACCGGAGCATATAAGT	1823			
QY	684	GGGACTATCTTCTTAACCCCGAGACCAACCGCAAGCAGCACCTATCGCTCTGC	743	DB	1791	CGCACGCTGGACCGCTGGCGCATGGCTTCCCAACTGGCAGAGAAAGATATGTGCT	1850			
DB	750	GGSTCTGATCTTGTGAACCCGGAGAGTGTATACGGCACCTGTATCOGCTGAGAACCG	809	QY	1824	ATCGGCTAGAGAAATGCTGTAGACCGGGGCGAGCTCTGTGCTCTTGGCACCAAGAT	1883			
QY	744	GGGTGATATCTGGTGAAGTTTGGCGAATGGCAATGATGATGACGAAGAACCGTGCT	803	DB	1851	GAAACCGGCAAGATGCTCTGCTGATCTGTGGCAAGCTGATGGCTTAAACCGCGGAAAT	1910			
DB	810	ACAGCAGGTACTTGAACCTTGGCCGCTATGGCATGGAACCGCAAGAAACCGACGCT	869	QY	1884	GACTCTTTGTAGCGGTATCGGTACTGGGCAACCACTAGCAGCTGGTTCGACGATGG	1943			
QY	804	CATAGCGGTGGACACCTCTCGGAANAACCATGGTGTGCGCATGGGAGAAATATGT	863	DB	1911	GACCGTCTGTCTGGCGGTATGCGGCTATGGCACCACCTATGTGGTGGCAACCAACG	1970			
DB	870	CACGCTGGCGGCGCAACCGCTGGATGTATGACGAGTATGGCAATGCTCTGCGT---T	926	QY	1944	AGTGTTFACAAATAGCGGGTCACTATCCAAATGACTCTTTGTAAACCTGCTAGACT	2003			
QY	864	GGGCGTGGACGCTTGGCGGCGGATGTGAAGAAATGACGCTGGGTGGAANAACCTA	923	DB	1971	CGTATTCAGCTTGTGAAGCGCATTTGACCAACGACTTTTGTGAACTTGACCTAT	2030			
DB	927	AGCCCTTGACCAAAAGCTCTGACGTGTGAACACAGGCTTAGTGTGGGCGAACCCCA	986	QY	2004	CMACACTAAATGSGCAGCAGGATGATATGACAAAGATTTTGAAGCGAGACTTCA	2063			
QY	924	CGGCAACGACAGCGGTGGCATTACCATCAGTGGACTAGAAGCGCTTCGACCAAGAC	983	DB	2031	GGGGAACACTGG-----AAGCGGTAGTGTAGCAAGCGCTAGCAAAATCGCGACGCA	2084			
DB	987	CATCGACGGCAGGCAAGCAACCGCGGCTACGCTGTATGAGGTGCTTGGACACCA	1046	QY	2064	AATCGGCAAGTAAATGTGGATGGCACCGGCTAGACTGATCTTCGGATCCCAATTCGA	2123			
QY	984	GCGTACTCATTTGAGCAATATCTTTTGAAGAGCTCTTTGGTACAGGTGGGAGCTTAC	1043	DB	2085	GACCGGTGCGGTGAAGTGTGACCGGCTCTCGCGGGTGGATCTGTGATTTTGTCTCAACTGCT	2144			
DB	1047	CCCAACCAATTCGATATGGGCTATTTTCGACCTGCTTGGCTGCAATATGSGAATCGA	1106	QY	2124	GCTTAAGAGCTCGCAGAGAGCTGACGCTGTGCGATCTCTGAAGAAAGATTTGTTAAGA	2183			
QY	1044	CAAAATGCCAGCTGAGGCTTATCAGTGGAAACCAAAAGACGCTGGCGGGCTGGCAACT	1103	DB	2145	ACTGGCGCTCTTAGCGCAAGTGTACGCGAGCAGTAAACGGCGAGATGTGTCAGAGA	2204			
DB	1107	AAGAGTCTTCGCGGTGCCACCAATTTGGGAAACCGGATTTGACATCAAAAGGAAACAGCC	1166	QY	2184	TTTTTGAGGCGCTGGGCGCAAGTAAATGACCTGGACCGGTTTGA	2228			
QY	1104	ACGGATGTCACATGATCCAGCAAGTGGCAGCGTCCATTTATGCTCACTACGAGCTGGC	1163	DB	2205	CTTCTGCGCGCTTGGACCAAGTGTATGACCGCCAGCTTTTGA	2249			

RESULT 15	
AB553944	
ID	AB553944 standard; DNA; 2262 BP.
AC	AB553944;
DT	22-NOV-2002 (first entry)
XX	
XX	DNA encoding Alcaligenes catalase protein 64C2A2.
XX	
XX	Gene; s8: marine bacteria; catalase; hydrogen peroxide; enzyme.
KW	
XX	
OS	Alcaligenes aquamarinus.
XX	
XX	Key Location/Qualifiers
FF	CDS 1..2262
FT	/*tag= a
FT	/product= "Catalase protein"
XX	
PN	US2002102660-A1.
XX	
PD	01-AUG-2002.
XX	
PD	19-JUN-2001; 2001US-0884889.
XX	
PF	03-JUL-1996; 96US-0674887.
PN	16-OCT-1997; 97US-0951844.
XX	
XX	95-OCT-1995; 95US-0412347.
XX	
PA	(ROBE/) ROBERTSON D E.
PA	(SANY/) SANYAL I.
PA	(ADHI/) ADHIKARI R.
XX	
XX	Robertson DE, Sanyal I, Adhikari R;
PI	WPI: 2002-690613/74.
XX	P-FSDb; ABC35040.
XX	
XX	New purified catalase polypeptide useful for catalyzing the breakdown
PT	of hydrogen peroxide, for modifying small molecules, and for generating
PT	antibodies which bind to the polypeptide
XX	
XX	Claim 1: Fig 5; 44pp: English.
PS	This invention relates to the DNA and protein sequences of a novel
CC	hydrolytic catalase polypeptide. The protein of the invention or its
CC	homologue is useful for modifying small molecules, for modifying
CC	the protein sequence of the invention is useful for modifying
CC	molecules, by mixing the protein or its fragments with a small molecule
CC	to produce a modified small molecule. The protein of the invention is
CC	also useful for catalyzing biochemical reactions, for hydrolysing
CC	glycosidic linkages and for generating antibodies which bind
CC	specifically to the protein. The nucleic acid sequences of the
CC	invention is useful as a probe to determine whether a biological sample,
CC	such as a soil sample, contains an organism having the nucleic acid or
CC	protein sequence of the invention, or for identifying an
CC	organism having the nucleic acid or protein sequence of the invention.
CC	Isolating related nucleic acids and proteins, and using the catalase
CC	protein of the invention is useful in immunoinfinity chromatography
CC	procedures to isolate or purify the protein or for detection of protein
CC	expressions in a biological sample. The cellulase protein of the
CC	invention is heat stable, is heat resistant, and is able to
CC	renature and regain activity after exposure to temperatures of from
CC	about 60 to 105 degrees. The present sequence represents the DNA
CC	sequence encoding the Alcaligenes aquamarinus catalase protein of the
XX	invention.
XX	
SD	Sequence 2262 BP; 553 A; 612 C; 663 G; 434 T; 0 other;



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: October 8, 2003, 07:56:28 | Search time 126.323 Seconds  
(without alignments)  
7819.780 Million cell updates/sec

Title: US-09-884-889-7  
Perfect score: 2238  
Sequence: 1 atggaaatcaccaacatc.....accggttgatcgaaataa 2238

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues  
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A-COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B-COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A-COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B-COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCOTS-COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than 1 to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2238	100.0	2238	2 US-08-674-887A-7	Sequence 7, Appl
2	2238	100.0	2238	3 US-08-951-844-7	Sequence 7, Appl
3	2238	100.0	2238	4 US-09-412-347-7	Sequence 7, Appl
4	844.4	37.7	2331	1 US-08-418-782-20	Sequence 20, Appl
5	844.4	37.7	2331	2 US-08-952-219-20	Sequence 20, Appl
6	844.4	37.7	4403765	3 US-09-103-840A-1	Sequence 1, Appl
7	844.4	37.7	4411529	3 US-08-418-782-1	Sequence 1, Appl
8	832.4	37.2	2235	1 US-08-228-662-1	Sequence 1, Appl
9	832.4	37.2	2235	2 US-08-852-219-1	Sequence 1, Appl
10	731.8	32.3	2235	1 US-08-418-782-2	Sequence 2, Appl
11	731.8	32.3	2235	2 US-08-418-782-2	Sequence 2, Appl
12	731.8	32.3	2221	1 US-08-228-662-2	Sequence 45, Appl
13	721.8	32.3	2221	2 US-08-852-219-2	Sequence 45, Appl
14	721.8	32.3	4795	2 US-08-952-219-2	Sequence 45, Appl
15	721.8	32.3	4795	3 US-08-952-219-2	Sequence 45, Appl
16	716.6	32.0	2181	4 US-08-032-614A-5	Sequence 8, Appl
17	695.2	30.0	2181	4 US-09-328-352-3989	Sequence 3989, Ap
18	672.2	30.0	2262	2 US-08-674-887A-5	Sequence 5, Appl
19	672.2	30.0	2262	3 US-08-951-844-5	Sequence 5, Appl
20	672.2	30.0	2262	4 US-09-412-347-5	Sequence 5, Appl
21	251.8	11.3	620	2 US-08-757-653-145	Sequence 145, App
22	251.8	11.3	620	2 US-08-520-946-145	Sequence 149, App
23	251.8	11.3	620	4 US-08-520-946-145	Sequence 149, App
24	251.8	11.3	620	2 US-08-757-653-145	Sequence 145, App
25	250.2	11.2	620	2 US-08-757-653-146	Sequence 146, App
26	250.2	11.2	620	2 US-08-757-653-147	Sequence 147, App
27	250.2	11.2	620	2 US-08-757-653-147	Sequence 147, App

Sequence 150, App  
Sequence 143, App  
Sequence 146, App  
Sequence 147, App  
Sequence 150, App  
Sequence 144, App  
Sequence 148, App  
Sequence 149, App  
Sequence 148, App  
Sequence 149, App  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-08-674-887A-7  
US-08-674-887A-7 Application US/08674887A  
Patent No. 5939300  
GENERAL INFORMATION:  
APPLICANT: Robertson, Dan E.  
INVENTOR: Subramanian, Indrajit  
APPLICANT: Subramanian, Indrajit  
TITLE OF INVENTION: CATALASES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: SEQESQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/674.887A  
FILING DATE: 03-JUL-1996  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: PATTON NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09015/002001  
TELEPHONE: 619/578-5070  
TELEFAX: 619/578-5099  
INFORMATION FOR SEQ ID NO:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2238 base pairs  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...2235  
US-08-674-887A-7

Query Match 100.0% Score 2238; DB 2; Length 2238;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2238; Conservative 0; Mismatches 0; Gaps 0;

QY 1 ATGGAAATCAACAACTCTAGATCTTCGTGATATACAAACAACTCTGGGAAATGC 60  
DB 1 ATGGAAATCAACAACTCTAGATCTTCGTGATATACAAACAACTCTGGGAAATGC 60  
QY 61 CATTATACCGAGGTTCGCTTATAGCAAGTGCAGGTGGCGACCAAAACAGGATTCG 120  
DB 61 CATTATACCGAGGTTCGCTTATAGCAAGTGCAGGTGGCGACCAAAACAGGATTCG 120  
QY 121 TGGCCCAACATCTCAACCTGGCATCTTACGCCCAACATTCATCGTATCGCAACAA 180  
DB 121 TGGCCCAACATCTCAACCTGGCATCTTACGCCCAACATTCATCGTATCGCAACAA 180  
QY 181 GACCGGATTTGATATGCGGCAAGCTTATAGCAAGTGCAGGTGGCGACCAAAAC 240  
DB 181 GACCGGATTTGATATGCGGCAAGCTTATAGCAAGTGCAGGTGGCGACCAAAAC 240  
QY 241 GACTTGGCAGCGCTATGACAGATTCACAGATTCACAGATTCGCGGACCAAGATTCAT 300  
DB 241 GACTTGGCAGCGCTATGACAGATTCACAGATTCACAGATTCGCGGACCAAGATTCAT 300  
QY 301 GACCTGCGAGCGCTATGACAGATTCACAGATTCACAGATTCGCGGACCAAGATTCAT 360  
DB 301 GACCTGCGAGCGCTATGACAGATTCACAGATTCACAGATTCGCGGACCAAGATTCAT 360  
QY 361 CTTGGTGGCGGTTCGCTGCGGCTGACAGATTCACAGATTCGCGGACCAAGATTCAT 420  
DB 361 CTTGGTGGCGGTTCGCTGCGGCTGACAGATTCACAGATTCGCGGACCAAGATTCAT 420  
QY 421 GCAATCTGCTATGAGAGCTGCTCTTGGCCCATCAACAACTACGCTGCAAA 480  
DB 421 GCAATCTGCTATGAGAGCTGCTCTTGGCCCATCAACAACTACGCTGCAAA 480  
QY 481 ATCTCTGGCGGATCAATGATATCAAGAAACCTAGCTCTGCAAACTATGGCTTT 540  
DB 481 ATCTCTGGCGGATCAATGATATCAAGAAACCTAGCTCTGCAAACTATGGCTTT 540  
QY 541 AAACCTTTTGGTTCGAGTGGCAGAGATGATGGAGCTTGAAGAGATGTATAC 600  
DB 541 AAACCTTTTGGTTCGAGTGGCAGAGATGATGGAGCTTGAAGAGATGTATAC 600  
QY 601 TGGGACAGAAACCGAATGGCTGGGAGAGATGATGGAGCTTGAAGAGATGTATAC 660  
DB 601 TGGGACAGAAACCGAATGGCTGGGAGAGATGATGGAGCTTGAAGAGATGTATAC 660  
QY 661 AATCCCTTGGAGCGCTACAAATGGGACTCATCTATGAAACCGGAGAGAGCTCGAA 720  
DB 661 AATCCCTTGGAGCGCTACAAATGGGACTCATCTATGAAACCGGAGAGAGCTCGAA 720  
QY 721 AAGCAGACCTTATCGCTGCTGGCGGTATTCGAGCTTTTGGCCGATGGCAATG 780  
DB 721 AAGCAGACCTTATCGCTGCTGGCGGTATTCGAGCTTTTGGCCGATGGCAATG 780  
QY 781 AATGACAGAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
DB 781 AATGACAGAAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
QY 841 GCTGCGATTCGAGAAATATGCTGCGGAGAGCTGCTGCGGAGAGATGAGAAATG 900  
DB 841 GCTGCGATTCGAGAAATATGCTGCGGAGAGCTGCTGCGGAGAGATGAGAAATG 900  
QY 901 AGCCTTGGGAGAGAAACCTAGCAGCACTGCGGAGAGCTGCTGCGGAGAGATGAGAAATG 960  
DB 901 AGCCTTGGGAGAGAAACCTAGCAGCACTGCGGAGAGCTGCTGCGGAGAGATGAGAAATG 960  
QY 961 CTAGAGGCGCTTGGACCAAGCCCTTACTGATGAGCAATGCTGCTGCTGCTGCTGCTGCT 1020  
DB 961 CTAGAGGCGCTTGGACCAAGCCCTTACTGATGAGCAATGCTGCTGCTGCTGCTGCTGCT 1020  
QY 1021 TTTGGTATGAGCGGAGCTTACCAAACTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
DB 1021 TTTGGTATGAGCGGAGCTTACCAAACTGCGAGCTTATCACTGCTGCTGCTGCTGCTGCTGCT 1080

QY 1081 GACGTGCGCGGCTTGGCACTATCCAGGATGACATGATCCAGAGATGCGAGCTCCA 1140  
DB 1081 GACGTGCGCGGCTTGGCACTATCCAGGATGACATGATCCAGAGATGCGAGCTCCA 1140  
QY 1141 TTTATGCTCTACTACGAGCTTGGCGCTTGGCGATGCGAGTGCAGCTGATTTACGAAA 1200  
DB 1141 TTTATGCTCTACTACGAGCTTGGCGCTTGGCGATGCGAGTGCAGCTGATTTACGAAA 1200  
QY 1201 CGGTACTATGAAAGCTTGTATGAGTTTGCAGATGCTTTCCGAAGCAATGTTACAA 1260  
DB 1201 CGGTACTATGAAAGCTTGTATGAGTTTGCAGATGCTTTCCGAAGCAATGTTACAA 1260  
QY 1261 ACACAGAGATATGCGGACCAAGAGTGCAGTACCTTGGACAGAGATGCTTACAACTG 1320  
DB 1261 ACACAGAGATATGCGGACCAAGAGTGCAGTACCTTGGACAGAGATGCTTACAACTG 1320  
QY 1321 CTATCTTGGCAAGACCTTATACAGATGATAGCAATGATAGCAATGATAGCAATGAT 1380  
DB 1321 CTATCTTGGCAAGACCTTATACAGATGATAGCAATGATAGCAATGATAGCAATGAT 1380  
QY 1381 GAAGCGCTTAAACCAAACTTCTGGATTCGGGATGAGGATGAGGATGAGGATGAGGAT 1440  
DB 1381 GAAGCGCTTAAACCAAACTTCTGGATTCGGGATGAGGATGAGGATGAGGATGAGGAT 1440  
QY 1441 GCATGGGCTTCTGCATCTTCTTGAATCTGGAATCGGATGAGGATGAGGATGAGGAT 1500  
DB 1441 GCATGGGCTTCTGCATCTTCTTGAATCTGGAATCGGATGAGGATGAGGATGAGGAT 1500  
QY 1501 GCTATACGACTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1560  
DB 1501 GCTATACGACTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1560  
QY 1561 GTACTTAAACCACTTGAAGATGATGAGAGTATGAGAGTATGAGAGTATGAGAGTATGAGAGT 1620  
DB 1561 GTACTTAAACCACTTGAAGATGATGAGAGTATGAGAGTATGAGAGTATGAGAGTATGAGAGT 1620  
QY 1621 GCATGATGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1680  
DB 1621 GCATGATGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1680  
QY 1681 AAGATGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1740  
DB 1681 AAGATGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1740  
QY 1741 GAGCAACCGATGCGAGCTTCTTGAAGCACTAGAGCACTAGAGCACTAGAGCACTAGAGCACT 1800  
DB 1741 GAGCAACCGATGCGAGCTTCTTGAAGCACTAGAGCACTAGAGCACTAGAGCACTAGAGCACT 1800  
QY 1801 FACATTAACCGGAGATTAAGATGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 1860  
DB 1801 FACATTAACCGGAGATTAAGATGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 1860  
QY 1861 CTGTGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1920  
DB 1861 CTGTGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1920  
QY 1921 AACTACGAGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1980  
DB 1921 AACTACGAGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCTTGGCGCT 1980  
QY 1981 TTTCTTTTAACTCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 2040  
DB 1981 TTTCTTTTAACTCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 2040  
QY 2041 GTTTTGAAGGAGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 2100  
DB 2041 GTTTTGAAGGAGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 2100  
QY 2101 CTGATCTTGGATCCATTTCCGAGCTGAGAGCTTCCGAGAGCTTCCGAGAGCTTCCGAGAGCTTCCGAGAGCT 2160  
DB 2101 CTGATCTTGGATCCATTTCCGAGCTGAGAGCTTCCGAGAGCTTCCGAGAGCTTCCGAGAGCTTCCGAGAGCT 2160  
QY 2161 TCTGAAGAAAGTTTGTAAAGCTTTGTGAAGCTTGTGAAGCTTGTGAAGCTTGTGAAGCTTGTGAAGCT 2220

```

Db 2161 TCGAGGAAAGTTGTTAAGAGTTTGTGAGGCGCTGGGCCAAAGTATGACCTGAGC 2220
Qy 2221 CGGTTTGATCTGAATAA 2238
Db 2221 CGGTTTGATCTGAATAA 2238

RESULT 2
US-09-951-844-7
US-09-951-844-7
Sequence 7, Application US/08951844
Patent No. 6074860
GENERAL INFORMATION:
APPLICANT: Robertson et al.
TITLE OF INVENTION: Catalases
NUMBER OF SEQUENCES: 8
COMPLIMENTARY SEQUENCES: 0
ADDRESSES: CARELLA, BYRNE, BALIN, GLIFILLAN,
ADDRESS2: CECHELI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER: VARIABLE FORM.
MODERN TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,844
FILING DATE: 435
PRIORITY DATE: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/674,887
FILING DATE: July 3, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Charles J. Herron
REGISTRATION NUMBER: 28,019
TELEPHONE: 201-994-1700
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2238 NUCLEOTIDES
TYPE: NUCLEIC ACID
SOURCE: HUMAN
TOPOLOGY: LINEAR
MOLECULE TYPE: DNA
US-08-951-844-7

Query Match 100.0%; Score 2238; DB 3; Length 2238;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGGAAATACACAACTACAGGATCTCTACGATATACAACTGSCGGAAATATGC 60
Db 1 ATGGAAATACACAACTACAGGATCTCTACGATATACAACTGSCGGAAATATGC 60
Qy 61 CTTTATCCCGAGGTTCGCTTAGCAAGTGCAGGTGGCGGCACCAAAACAGGATGG 120
Db 61 CTTTATCCCGAGGTTCGCTTAGCAAGTGCAGGTGGCGGCACCAAAACAGGATGG 120
Qy 121 TGGCCCCACATGCTCAACCTCGCATCTTAGCGCCACATCTCATCGCTATCGGACCCAAAC 180
Db 121 TGGCCCCACATGCTCAACCTCGCATCTTAGCGCCACATCTCATCGCTATCGGACCCAAAC 180
Qy 181 GACCCGGATTTGACTATGCGGAAGGTTTAAAGAGCTAGATCTGGAGCGGTTAAAGAG 240
Db 181 GACCCGGATTTGACTATGCGGAAGGTTTAAAGAGCTAGATCTGGAGCGGTTAAAGAG 240
Qy 241 GACCTGGCAGCGCTATGACAGATTCACAGGACTGTGGCGCAGCAGATTAAGCTCATTTAT 300

```

```

Db 241 GACCTGGCAGCGCTATGACAGATTCACAGGACTGTGGCGCAGCAGATTAAGCTCATTTAT 300
Qy 301 GGGCCCTCTTTATACGATGGCGTGGCACAGCGCGGCACCTACCGTATCGGTATGGC 360
Db 301 GGGCCCTCTTTATACGATGGCGTGGCACAGCGCGGCACCTACCGTATCGGTATGGC 360
Qy 361 CGTGGTGGCGGTGGCGCTTCAGAGCGCTTCGCGCTCTCATAGCTGGCAGACAT 420
Db 361 CGTGGTGGCGGTGGCGCTTCGCGCTTCAGAGCGCTTCGCGCTCTCATAGCTGGCAGACAT 420
Qy 421 GCCATCTGATTAAGCAGCTGCTCTTTTGGCCCATCAACAAAAATACGGTCGAAA 480
Db 421 GCCATCTGATTAAGCAGCTGCTCTTTTGGCCCATCAACAAAAATACGGTCGAAA 480
Qy 481 ATCTCTGGCGGATCTAATGACTACACAGAACCTAGCTCTGGAACATATGGCTTT 540
Db 481 ATCTCTGGCGGATCTAATGACTACACAGAACCTAGCTCTGGAACATATGGCTTT 540
Qy 541 AAAATCTTGGTTTTCAGGTGGCAGACAGATGTATGGAGCCTCAAGAGATGTATAC 600
Db 541 AAAATCTTGGTTTTCAGGTGGCAGACAGATGTATGGAGCCTCAAGAGATGTATAC 600
Qy 601 TGGGAGCAGAACCGATGGCTTGGGAGACAGCGCTATGAAGTGAACGAGCTCGAA 660
Db 601 TGGGAGCAGAACCGATGGCTTGGGAGACAGCGCTATGAAGTGAACGAGCTCGAA 660
Qy 661 AATCCCTTGGGAGCGGTACAAATGGGACTCATCTATGTAACCCCGAAGCAACGCG 720
Db 661 AATCCCTTGGGAGCGGTACAAATGGGACTCATCTATGTAACCCCGAAGCAACGCG 720
Qy 721 AAGCAGACCCCTATCGCTGCTGGCGTGATATCGTGAGACTTTTGGCCGATGGCAATG 780
Db 721 AAGCAGACCCCTATCGCTGCTGGCGTGATATCGTGAGACTTTTGGCCGATGGCAATG 780
Qy 781 AATGACAGAAACCGCTGCTCTATAGCGGTGGACACACCTTCGAAAAAACCCATGT 840
Db 781 AATGACAGAAACCGCTGCTCTATAGCGGTGGACACACCTTCGAAAAAACCCATGT 840
Qy 841 GCTGCCGATGCGGAAATATGTGGCGGAGAGCTCGCGCGCAGGATATTGAAGAAATG 900
Db 841 GCTGCCGATGCGGAAATATGTGGCGGAGAGCTCGCGCGCAGGATATTGAAGAAATG 900
Qy 901 AGCTGGGTGGAAACACCTACGCGCACCGAGCGTGGGATACATCACCAGTGA 960
Db 901 AGCTGGGTGGAAACACCTACGCGCACCGAGCGTGGGATACATCACCAGTGA 960
Qy 961 CTAGAGGCGCTGCGACAGACCCCTACTCATTTGGAGCAATACTTTTGAACACTC 1020
Db 961 CTAGAGGCGCTGCGACAGACCCCTACTCATTTGGAGCAATACTTTTGAACACTC 1020
Qy 1021 TTTGTTACAGATGGGATTTACAAAAAGTTCGAGCTGAGCTTACGTGGAAACCAAA 1080
Db 1021 TTTGTTACAGATGGGATTTACAAAAAGTTCGAGCTTACGTGGAAACCAAA 1080
Qy 1081 GACGCTGGGGGCTGCGACCATCGATGATGCGGATGCGCAGGATGCGCAGCTCA 1140
Db 1081 GACGCTGGGGGCTGCGACCATCGATGATGCGGATGCGCAGGATGCGCAGCTCA 1140
Qy 1141 TTTATGCTACTACGAGACCTGGCGCTGGGATGGGATGGGATGGGATGGGATGGG 1200
Db 1141 TTTATGCTACTACGAGACCTGGCGCTGGGATGGGATGGGATGGGATGGGATGGG 1200
Qy 1201 CGGTACTATGAACACCTGATGATGTTTGCAGATGCTTCCCGAAAGCATGGTACAACTC 1260
Db 1201 CGGTACTATGAACACCTGATGATGTTTGCAGATGCTTCCCGAAAGCATGGTACAACTC 1260
Qy 1261 ACACACAGATATGGGACCAAGGTGGCTTACCTGGGACAGAAAGTGGCTCAGGAGAC 1320
Db 1261 ACACACAGATATGGGACCAAGGTGGCTTACCTGGGACAGAAAGTGGCTCAGGAGAC 1320
Qy 1321 CTCATCTGGCAGACCTTACAGATGATGAGCATCTCTCTGTAGACCAAAACGATAT 1380
Db 1321 CTCATCTGGCAGACCTTACAGATGATGAGCATCTCTCTGTAGACCAAAACGATAT 1380

```





481. ATCTCTGGCGGAGCTAATGTAATCTACAGCAAGCTGAGCAATGAGCTTT 540  
 Db ATCTCTGGCGGAGCTAATGTAATCTACAGCAAGCTGAGCAATGAGCTTT 540  
 541. AATACCTTTTGGTTTTCAGAGTGGCAGACAGATGATGAGGAGCTGAGAGATGATATAC 600  
 Qy ATCTCTGGCGGAGCTAATGTAATCTACAGCAAGCTGAGCAATGAGCTTT 540  
 Db ATCTCTGGCGGAGCTAATGTAATCTACAGCAAGCTGAGCAATGAGCTTT 540  
 541. AATACCTTTTGGTTTTCAGAGTGGCAGACAGATGATGAGGAGCTGAGAGATGATATAC 600  
 Qy ATCTCTGGCGGAGCTAATGTAATCTACAGCAAGCTGAGCAATGAGCTTT 540  
 Db ATCTCTGGCGGAGCTAATGTAATCTACAGCAAGCTGAGCAATGAGCTTT 540  
 601. TGGGAGCAGCAAGCAAGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 660  
 Qy TGGGAGCAGCAAGCAAGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 660  
 Db TGGGAGCAGCAAGCAAGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 660  
 601. TGGGAGCAGCAAGCAAGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 660  
 Qy TGGGAGCAGCAAGCAAGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 660  
 Db TGGGAGCAGCAAGCAAGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 660  
 661. AATCCCTGGGAGCGCTCAAAATGGGAGCTCACTATGTAATACCCGAGGAGCCAAACGGC 720  
 Qy AATCCCTGGGAGCGCTCAAAATGGGAGCTCACTATGTAATACCCGAGGAGCCAAACGGC 720  
 Db AATCCCTGGGAGCGCTCAAAATGGGAGCTCACTATGTAATACCCGAGGAGCCAAACGGC 720  
 661. AATCCCTGGGAGCGCTCAAAATGGGAGCTCACTATGTAATACCCGAGGAGCCAAACGGC 720  
 Qy AATCCCTGGGAGCGCTCAAAATGGGAGCTCACTATGTAATACCCGAGGAGCCAAACGGC 720  
 Db AATCCCTGGGAGCGCTCAAAATGGGAGCTCACTATGTAATACCCGAGGAGCCAAACGGC 720  
 721. AAGCCAGCACTGATGCTGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 780  
 Qy AAGCCAGCACTGATGCTGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 780  
 Db AAGCCAGCACTGATGCTGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 780  
 721. AAGCCAGCACTGATGCTGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 780  
 Qy AAGCCAGCACTGATGCTGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 780  
 Db AAGCCAGCACTGATGCTGCTGAGCAGCAAGCTGAGCAAGCTGAGCAAGCTGAGCA 780  
 781. AATGACGAGAAACCTGGCTCTCATAGCGGGTGGACACACTTCGGAAAACCAATGGT 840  
 Qy AATGACGAGAAACCTGGCTCTCATAGCGGGTGGACACACTTCGGAAAACCAATGGT 840  
 Db AATGACGAGAAACCTGGCTCTCATAGCGGGTGGACACACTTCGGAAAACCAATGGT 840  
 781. AATGACGAGAAACCTGGCTCTCATAGCGGGTGGACACACTTCGGAAAACCAATGGT 840  
 Qy AATGACGAGAAACCTGGCTCTCATAGCGGGTGGACACACTTCGGAAAACCAATGGT 840  
 Db AATGACGAGAAACCTGGCTCTCATAGCGGGTGGACACACTTCGGAAAACCAATGGT 840  
 841. GCTGCGGATGCGGAGAAATATGCTGGCGGAGCTGCGCGCGGAGGATTCGAGAAATG 900  
 Qy GCTGCGGATGCGGAGAAATATGCTGGCGGAGCTGCGCGCGGAGGATTCGAGAAATG 900  
 Db GCTGCGGATGCGGAGAAATATGCTGGCGGAGCTGCGCGCGGAGGATTCGAGAAATG 900  
 841. GCTGCGGATGCGGAGAAATATGCTGGCGGAGCTGCGCGCGGAGGATTCGAGAAATG 900  
 Qy GCTGCGGATGCGGAGAAATATGCTGGCGGAGCTGCGCGCGGAGGATTCGAGAAATG 900  
 Db GCTGCGGATGCGGAGAAATATGCTGGCGGAGCTGCGCGCGGAGGATTCGAGAAATG 900  
 901. ASCTGGGGTGGAAAACACTTACGCGCAGCGACACGCTGGCGGATACCATACCGATGGA 960  
 Qy ASCTGGGGTGGAAAACACTTACGCGCAGCGACACGCTGGCGGATACCATACCGATGGA 960  
 Db ASCTGGGGTGGAAAACACTTACGCGCAGCGACACGCTGGCGGATACCATACCGATGGA 960  
 901. ASCTGGGGTGGAAAACACTTACGCGCAGCGACACGCTGGCGGATACCATACCGATGGA 960  
 Qy ASCTGGGGTGGAAAACACTTACGCGCAGCGACACGCTGGCGGATACCATACCGATGGA 960  
 Db ASCTGGGGTGGAAAACACTTACGCGCAGCGACACGCTGGCGGATACCATACCGATGGA 960  
 961. CTAGAGCGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1020  
 Qy CTAGAGCGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1020  
 Db CTAGAGCGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1020  
 961. CTAGAGCGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1020  
 Qy CTAGAGCGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1020  
 Db CTAGAGCGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1020  
 1021. TTGGTTACGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA 1080  
 Qy TTGGTTACGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA 1080  
 Db TTGGTTACGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA 1080  
 1021. TTGGTTACGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA 1080  
 Qy TTGGTTACGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA 1080  
 Db TTGGTTACGAGTGGAGCTTACCAAAAGTCCAGCTGGAGCTTATCAGTGGAAACCAAAA 1080  
 1081. GACGTGCGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1140  
 Qy GACGTGCGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1140  
 Db GACGTGCGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1140  
 1081. GACGTGCGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1140  
 Qy GACGTGCGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1140  
 Db GACGTGCGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1140  
 1141. TTATGCTCCTACTAGGAGCTGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1200  
 Qy TTATGCTCCTACTAGGAGCTGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1200  
 Db TTATGCTCCTACTAGGAGCTGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1200  
 1141. TTATGCTCCTACTAGGAGCTGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1200  
 Qy TTATGCTCCTACTAGGAGCTGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1200  
 Db TTATGCTCCTACTAGGAGCTGGCGCTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1200  
 1201. CGGTACTATGAAACCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 Qy CGGTACTATGAAACCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 Db CGGTACTATGAAACCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 1201. CGGTACTATGAAACCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 Qy CGGTACTATGAAACCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 Db CGGTACTATGAAACCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 1261. ACACAGAGATGAGGAGCAAGCTGGCTACCTGGGAGCAGAGTGGCTCAGGAGAC 1320  
 Qy ACACAGAGATGAGGAGCAAGCTGGCTACCTGGGAGCAGAGTGGCTCAGGAGAC 1320  
 Db ACACAGAGATGAGGAGCAAGCTGGCTACCTGGGAGCAGAGTGGCTCAGGAGAC 1320  
 1261. ACACAGAGATGAGGAGCAAGCTGGCTACCTGGGAGCAGAGTGGCTCAGGAGAC 1320  
 Qy ACACAGAGATGAGGAGCAAGCTGGCTACCTGGGAGCAGAGTGGCTCAGGAGAC 1320  
 Db ACACAGAGATGAGGAGCAAGCTGGCTACCTGGGAGCAGAGTGGCTCAGGAGAC 1320  
 1321. CTGATCTGGCAGAGCCCTATACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 Qy CTGATCTGGCAGAGCCCTATACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 Db CTGATCTGGCAGAGCCCTATACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 1321. CTGATCTGGCAGAGCCCTATACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 Qy CTGATCTGGCAGAGCCCTATACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 Db CTGATCTGGCAGAGCCCTATACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
 1381. GAGGCGCTTAAAGCCCAAAATGCTGGATGAGCAGCTGAGCAGTGAAGTGGTGAAGCAG 1440  
 Qy GAGGCGCTTAAAGCCCAAAATGCTGGATGAGCAGCTGAGCAGTGAAGTGGTGAAGCAG 1440  
 Db GAGGCGCTTAAAGCCCAAAATGCTGGATGAGCAGCTGAGCAGTGAAGTGGTGAAGCAG 1440  
 1381. GAGGCGCTTAAAGCCCAAAATGCTGGATGAGCAGCTGAGCAGTGAAGTGGTGAAGCAG 1440  
 Qy GAGGCGCTTAAAGCCCAAAATGCTGGATGAGCAGCTGAGCAGTGAAGTGGTGAAGCAG 1440  
 Db GAGGCGCTTAAAGCCCAAAATGCTGGATGAGCAGCTGAGCAGTGAAGTGGTGAAGCAG 1440  
 1441. GATGGGCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCT 1500  
 Qy GATGGGCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCT 1500  
 Db GATGGGCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCT 1500  
 1441. GATGGGCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCT 1500  
 Qy GATGGGCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCT 1500  
 Db GATGGGCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCTTCTGATCT 1500  
 1501. CGTATGAGCTGGGCGCCCAAAAAGCTGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1560  
 Qy CGTATGAGCTGGGCGCCCAAAAAGCTGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1560  
 Db CGTATGAGCTGGGCGCCCAAAAAGCTGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1560  
 1501. CGTATGAGCTGGGCGCCCAAAAAGCTGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1560  
 Qy CGTATGAGCTGGGCGCCCAAAAAGCTGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1560  
 Db CGTATGAGCTGGGCGCCCAAAAAGCTGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1560  
 1561. GTACTCAAAACACTAGAGGTATCCAGGAGCTTTTAAACGAGCGCAATGATGATGATGATGATGAT 1620  
 Qy GTACTCAAAACACTAGAGGTATCCAGGAGCTTTTAAACGAGCGCAATGATGATGATGATGATGAT 1620

## RESULT 4.

US-08-418-782-20

: Sequence 20, Application US/08418782

: Patent No. 5658733

: GENERAL INFORMATION:

: APPLICANT: Cockerill, Franklin R.

: APPLICANT: Line, Bruce C.

: APPLICANT: Line, Bruce C.

: TITLE OF INVENTION: Detection of Isoniazid Resistant Strains

: TITLE OF INVENTION: of M. Tuberculosis

: NUMBER OF SEQUENCES: 22

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Schwegman, Lundberg &amp; Moessner

: STREET: 3500 IDS Center

: CITY: Minneapolis

: STATE: MN

: COUNTRY: USA

: ZIP: 55402

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: CURRENT APPLICATION DATE: Patent Release #1.0, Version #1.25

: APPLICATION NUMBER: US/08/418,782

PILING DATE: 37.78; Score 844.4; DB 1; Length 2331;  
 Best Local Similarity 63.0%; Pred No. 5, 4e-51;  
 Matches 1990; Conservative 0%; Mismatches 801; Indels 15; Gaps 5;  
 30 TAGCTATACACAAATGCGGGAAATGCCCTTTTACCGAGAGTTCGCTTACCAAG 89  
 93 TACAGAAACCAACCCGAGGCGGTAGCACMGCGCTGTCCGCTGGGTCAATAGAATA 152  
 90 TGCAGTGGCGGCCACCAAAACAGGATGGTGGCGCAACATCTCAACTCGGCATCTT 149  
 153 CCGCTGTGAGGGGGCGGAAACAGCATGGTGGCCACCGGCTCAATCTGAAGTACT 212  
 150 AGCCCAACATATGCTGTACGACCAACAGACGGATTTTGAATATGCCAGAGTT 209  
 213 GCACCAAAACCGGCTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 272  
 210 TAAGAAGCTAGATCTGCGACGGGTAAAGAGGAGCTGGCAGCGCTATGACAGATTCA 269  
 273 CGGRCATCGAGCTGAGCGGCTGACGGCGGATCGAGGAGATGAGTACCACTCGCA 332  
 270 GAGTGGTGGCCGACGATACGCTGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 329  
 333 CGCGTGGTGGCCGCGACTAGCGGCACTAGCGGCGGCTGTATATCGGATGGCTGGCA 392  
 390 CAGCGCGGCGACTACCGTATGGTGTAGTGGCTGGTGGTGGTGGTGGTGGTGGTGG 389  
 393 CGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 452  
 390 GTTGGCGGCTGCTATAGTGGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCG 449  
 453 GTTGGCGGCGGCTTACAGCTGGCGGAGATGCGGAGATGCGGAGATGCGGAGATG 512  
 450 TTGGCGGCTACAAACAAATACGCTGGAATATCTGCGGCGGAGATGCGGAGATGCG 509  
 513 GTGGCGGCTCAGAGAGATGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCG 572  
 510 AGGAACCTAGCTTGGAAATCTATGGGCTTTAAACTTTTGGTGGTGGTGGTGGTGG 569  
 573 CGGCACTGCGGCTGGAATCGATGGGCTTCAAGAGTGTGGGCTTGGGCTTGGGCT 632  
 570 AGATGTATGGGAGCTGAAGAAGATGATATCTGGGCGGAGATGCGGAGATGCGGAG 629  
 633 CAGCGAGTGGGAGGCTTGGGAGATGATATCTGGGCGGAGATGCGGAGATGCGGAG 689  
 630 CAGCGGCTATGAGGATGCGGAGATGATATCTGGGCGGAGATGCGGAGATGCGGAG 749  
 690 TGAGCGCTTACAGCGGTACGCGGATGCGGAGATGCGGAGATGCGGAGATGCGGAG 689  
 690 CATCTATGAAACCCGAGAGATGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCG 749  
 750 GATCTAGCTGAACCCGAGAGGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCG 809

Query Match 37.78; Score 844.4; DB 1; Length 2331;  
 Best Local Similarity 63.0%; Pred No. 5, 4e-51;  
 Matches 1990; Conservative 0%; Mismatches 801; Indels 15; Gaps 5;  
 30 TAGCTATACACAAATGCGGGAAATGCCCTTTTACCGAGAGTTCGCTTACCAAG 89  
 93 TACAGAAACCAACCCGAGGCGGTAGCACMGCGCTGTCCGCTGGGTCAATAGAATA 152  
 90 TGCAGTGGCGGCCACCAAAACAGGATGGTGGCGCAACATCTCAACTCGGCATCTT 149  
 153 CCGCTGTGAGGGGGCGGAAACAGCATGGTGGCCACCGGCTCAATCTGAAGTACT 212  
 150 AGCCCAACATATGCTGTACGACCAACAGACGGATTTTGAATATGCCAGAGTT 209  
 213 GCACCAAAACCGGCTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 272  
 210 TAAGAAGCTAGATCTGCGACGGGTAAAGAGGAGCTGGCAGCGCTATGACAGATTCA 269  
 273 CGGRCATCGAGCTGAGCGGCTGACGGCGGATCGAGGAGATGAGTACCACTCGCA 332  
 270 GAGTGGTGGCCGACGATACGCTGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 329  
 333 CGCGTGGTGGCCGCGACTAGCGGCACTAGCGGCGGCTGTATATCGGATGGCTGGCA 392  
 390 CAGCGCGGCGACTACCGTATGGTGTAGTGGCTGGTGGTGGTGGTGGTGGTGGTGG 389  
 393 CGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 452  
 390 GTTGGCGGCTGCTATAGTGGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCG 449  
 453 GTTGGCGGCGGCTTACAGCTGGCGGAGATGCGGAGATGCGGAGATGCGGAGATG 512  
 450 TTGGCGGCTACAAACAAATACGCTGGAATATCTGCGGCGGAGATGCGGAGATGCG 509  
 513 GTGGCGGCTCAGAGAGATGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCG 572  
 510 AGGAACCTAGCTTGGAAATCTATGGGCTTTAAACTTTTGGTGGTGGTGGTGGTGG 569  
 573 CGGCACTGCGGCTGGAATCGATGGGCTTCAAGAGTGTGGGCTTGGGCTTGGGCT 632  
 570 AGATGTATGGGAGCTGAAGAAGATGATATCTGGGCGGAGATGCGGAGATGCGGAG 629  
 633 CAGCGAGTGGGAGGCTTGGGAGATGATATCTGGGCGGAGATGCGGAGATGCGGAG 689  
 630 CAGCGGCTATGAGGATGCGGAGATGATATCTGGGCGGAGATGCGGAGATGCGGAG 749  
 690 TGAGCGCTTACAGCGGTACGCGGATGCGGAGATGCGGAGATGCGGAGATGCGGAG 689  
 690 CATCTATGAAACCCGAGAGATGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCG 749  
 750 GATCTAGCTGAACCCGAGAGGCGGAGATGCGGAGATGCGGAGATGCGGAGATGCG 809

750 TATCTGTGAGACTTTTGGCCGAATGGAATGAATGATGATGATGATGATGATGATGAT 809  
 810 CATCTGCGGAGACTTTTGGCCGAATGGAATGAATGATGATGATGATGATGATGATGAT 869  
 810 GGGTGGACACACTTTTGGCCGAATGGAATGATGATGATGATGATGATGATGATGATGAT 869  
 870 CGGCGGTGACACTTTTGGCCGAATGGAATGATGATGATGATGATGATGATGATGATGAT 869  
 870 AGAGCTGCGCGCGCGAGTATTGAAGAAATGAGCTGGGCTGGGCTGGGCTGGGCTGGG 926  
 927 GCGGCGGCGCGCGAGTATTGAAGAAATGAGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 926  
 930 GCGACACGCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 986  
 987 GCGACGCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 989  
 990 TCAATGAGCAATTAATTTTTTGAAGAAATGAGCTGGGCTGGGCTGGGCTGGGCTGGG 1046  
 1047 GAGTGGGCAACAGTTCCTGGAGATCTGTACGCGTACGAGTGGGAGCTGAGGAGAG 1049  
 1050 TCCAGCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1106  
 1107 CCGCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1109  
 1110 TGCACATGATCCCAAGTGGCAGCTTCAATTTATGCTCACTAGGAGCTGGGCTGGGCTGG 1166  
 1167 CCGTGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1169  
 1170 CATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1223  
 1224 GGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1229  
 1230 AGATGCTTTCGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1283  
 1284 CGAGGATGCTGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1289  
 1290 CTAGCTGGGAGCAGAGTGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1343  
 1344 ATGCTTGGGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1403  
 1350 AGGCACTCTCTTGTAGAGAAAGATTTGAAGGCTTAAAGCCAAATCTTGGAGCTTTCG 1409  
 1404 CAGGCACTCTCTTGTAGAGAAAGATTTGAAGGCTTAAAGCCAAATCTTGGAGCTTTCG 1463  
 1410 GGGACTGACGATGAGGAGCTTGTAGAGGCTTAAAGCCAAATCTTGGAGCTTTCG 1469  
 1464 GGGTGGCTTGT 1523  
 1470 CTCTGACAGCGGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1523  
 1524 TACGACAGAGCGGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1583  
 1530 GGAAGTAAACAACTCTAAGGCTTAAAGGCTTAAAGGCTTAAAGGCTTAAAGGCTTAAAGGCT 1586  
 1584 GAGGCTGACGAGCGGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1643  
 1587 GAGGCTGACGAGCGGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1646  
 1644 GAGGCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1703  
 1647 GCTGGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1706  
 1704 GCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1763  
 1707 GCTTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1766  
 1764 GCTTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 1823  
 1767 AGCATAGGAGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1826  
 1824 GCTGCTGGGAGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1883  
 1827 GCTGAGGAGT 1886

```

Db 1884 GCGGAGTACATGCTGCTGCAAGGGAACCTGCTTACGCTCAGTGCCTCCCTGAGTGC 1943
Qy 1887 TGCTTTGTAGAGGTATGCTGCTGACGACCACTACGACGCTTGGCAGCTGAGT 1946
Db 1944 GTTGCTGTAGTGGCTGCGCGTCTGCGCGCACTACAGCGCTTACCGCTGGCGGT 2003
Qy 1947 GTTTACAAATAGCGGGTCAAGTATCCATGATCTTCTTGTAAACCTGCTAGACCTCAA 2006
Db 2004 GTTACCGAGGCTCCAGTCACTGACGACCACTTCTGTCGACCTGCTGACATGGG 2063
Qy 2007 CACTTAACGCGCAGCAGCATGATCAGACAAATTTTGAAGCAGAGACTTCAAAC 2066
Db 2064 TATCACTCTGGAGCCCTCCGACGAGATGAGCGGACCTTACAGGCGCAGGATGCG 2120
Qy 2067 TGGCGAATGATGATGATGGGACCGGCTGAGCTGATCTCGGATTCGAATTCGAGCT 2126
Db 2121 TGGCAGTGAATGACACCGAGCGCTGACCTGCTGCTGCGCTCCACTCGAGTT 2180
Qy 2127 AAGAGCCCTCGCAGAGCTGACGGCTGTACAGATTCCTGAAGAAAGTTTGTAAAGATT 2186
Db 2181 GCGCGCGCTTGTGAGGCTCTATGGCGCCGATGACGGCGACGAAAGTTCTGCGCAGGACT 2240
Qy 2187 TGTGAGGCTGTGGCGCAAGTATGAGCTGTGACGCGGTTGATCTG 2232
Db 2241 GGTGCTGCTGGGCAAGCTGTATGACCTGACGACTGCTGACGTTGCGCTG 2286

```

## RESULT 5

```

US-08-852-219-20
; Sequence 20, Application US/08852219
; Patent No. 5922575
; GENE/SEQUENCE: HLA-B*57:01
; APPLICANT: Cockrell, Franklin R.
; APPLICANT: Kline, Bruce C.
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains
; TITLE OF INVENTION: Detection of M. Tuberculosis
; NUMBER OF SEQUENCES: 22
; CONGRESS ADDRESS: 22
; ADDRESS: 22
; STREET: 119 No. 5922575th Fourth Street, Ste. 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM: disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 07-May-1997
; CLASSIFICATION: US/08/852,219
; NAME: Cockrell, Franklin R.
; REGISTRATION NUMBER: 41,287
; REFERENCE/DOCKET NUMBER: 230.00010130
; TELEPHONE: 612-305-1226
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 20:
; SOURCE CHARACTERISTICS:
; LENGTH: 213
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 70..2289
; US-08-852-219-20

```

```

Query Watch 37.78; Score 844.4; DB 2; Length 2331;
Best Local Similarity 37.08; Pct. 5.4e-231;
Matches 1390; Conservative 801; Indels 15; Gaps 5;
Qy 30 TACGTATACGACAACTACGCGGAAATGCCCTTTTACGCGAGGTTTCCGTTAAGCAAG 89
Db 93 TACGAGTACGCGGACCAAAAGAGGATGTGTGGCCCAACATGCTCACTCGGCAATG 152
Qy 90 TCGAGTACGCGGACCAAAAGAGGATGTGTGGCCCAACATGCTCACTCGGCAATG 149
Db 153 CCCGCTCGAGGCGGAAACAGGAGTGTGTGCCCAACCGCTCACTATGGAAGTCT 202
Qy 150 AGCGCATATCATGCTATCGGACCCCAAAAGAGCGGATTTGACTATTCGCGGAGAGT 219
Db 213 GCACAAAACCGGCGCTCGCTGACCGATGATGGTGGCGGCTGCACTATTCGCGGAGGT 272
Qy 210 TAAAGAGCTAGATCTGCGAGCGGTATTAAGAGGCTCTGCGCGCTATACAGATCA 269
Db 273 CGCGACATCGAGCTTTCAGCGCTGACGCGGACATCAGAGGATGATGACCACTCGCA 332
Qy 270 GGAGCTGGTGGCGACGAGTATACGGTCTATATGGCCCTCTTTTATACGATGGCGTGA 329
Db 333 GCGCTGGTGGCGGCGCGGCTACGGCGCTACGGCGCTGTTTATTCGAGTGGCGTGA 392
Qy 330 GAGCGCGCTGACGCTACGCTGATGGGATGGGCTGTGTGGGCTGCGCTCGACGCG 389
Db 393 CGCTGCGCGGACCTTACGCGATCCACAGCGCGCGCGCGCGCGCGCGCGCGCGCTG 452
Qy 390 CTTGCGGCGCTCTCATATGCTGGCGAGCAATGCCAATCTGATAAAGCACTGCTTCT 449
Db 453 GTTCGCGCGGCTTAAACGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCT 512
Qy 450 TTGCGCGCATCAAGCAAAATACGCTCGAANAATCTCTGGCGGATCTGATGATGATG 509
Db 513 GTGGCGCTCAAGAGAGTACGCGCAAGAGCTCTCATGCGCGAGCTGATGTTTCTGCG 572
Qy 510 AGGAAACGTAGCTTGGAACTATGGGCTTAAAACTTTTGGTTTTCGAGGTGGCAGAG 569
Db 573 CGGCGCTGCGCGCTGGAATGATGGCTTCAAGAGCTTTCGGCTTCGCGCGCGGT 632
Qy 570 AGATGATGGGAGCTCGAAGAGTATACTTGGGCGAGCAGCAACGATGCTGCGGAG 629
Db 633 CGAGCAGTGGGAGCC--CGATGAGTCTATTGGGCGAAGGAGCACTGCTGCGCGA 689
Qy 630 CAAGCGCTATGAAGTGACCGAGAGCTCGAAATTCCTTGGGAGCGCTCAAAATGGAGCT 689
Db 690 TCGCGCTTACCGGCTTACGGGCTGAGAGACCGCTTGGCGGCGCTGAGATGGGCT 749
Qy 690 CATCTATGTAACCCCGAGAGCCACCGAGCAGCAGAGCTTATCGCTGCTGCGCGTGA 749
Db 750 GATCTACTGTAACCGGAGGCGCGGACGCGCAACCGCGACCCATTCGCGCGCGCTGCA 809
Qy 750 TATCTGCTGAGACTTTTGGCGCAATGCAATGATGAGAGAAACCGTGGCTTCTATAGC 809
Db 810 CATTCGCGAGCTTTCGCGGATGCGCTTGAACGAGCTGCGAAGCGCGGCTGATGCT 869
Qy 810 GGGTGAGACAGCTTTCGCGGAAACCGCTGCTGCGCGAGCGAGAAATGATGCTGCG 869
Db 870 CGGCGCTGACACTTTTGGGTAGACCCATGCGCG--CGCGCGCGCGCTGCTGCTGCGCG 926
Qy 870 AGAGCTGCGCGCGAGGATTTAGAGAAATGAGCTGAGGCTGGAAGAAACACTACGCGAC 929
Db 927 CGAAGCGGAGCTTTCGCGCTGAGGAGATGAGGCTTGGGCTTGGAGAGCTGATAGCGAC 986
Qy 930 CGGAGAGCTGCGGATACCTACAGTGGCTGAGAGGCGCTGAGAGAGAGCGCTGAC 989
Db 987 CGGAACCGGTAAAGACCGGATCACAGCGGATCGATGATGAGAGAGAGCGCGAC 1046
Qy 990 TCAATGGAGCAATAATTTTTTGAAGACTCTTTTGGTACGAGTGGGAGCTTACAAAG 1049
Db 1047 GAATGGGAGCAAGTCTTCTCGAGATCTCTGCTAGAGTGGGAGCTGAGAGAG 1106

```

QY 1050 TCCAGCTGAGGCTTATCATGTGGAAACAAAAGACGGCGCGGCTGGCAACCATACCGGA 1109  
 Db 1107 CCTCTGGCTTGGCATATACCCGACAGAGCGGGCGGTGGCGGCACCATCCCGGA 1166  
 QY 1110 TGCACATGATCCACAGTGGCCAGCTCATATCTCTACTACAGAGCTGGCGCTGG 1169  
 Db 1107 CCTGTTCG---GGGGCGAGGGCGCTCCCGGAGATGCTGGCCACTGACCTCTCGCTGG 1223  
 QY 1160 CAGGACCCCTGATACGAMAMATTTCTCGACGGTACTATGAACCCCTGATGATTTGC 1229  
 Db 1224 GGTGGATCGGATATATGGCGGATACAGCGTGGCTGGGACACCGGAGAAATGGC 1283  
 QY 1230 AGATGCTTTCCGCAAGCATGTACAACTGACACGACGATATGACGAGCAAGAGTGG 1289  
 Db 1284 CGACGATTTCCGCAAGCGCTGTACAGCTGATCCACGACGATGCTCGCTGGAC 1343  
 QY 1290 CTATCTGGACCAAGTGCCTCAGAGACGACTCATCTGCGACGACCTATACCAAGT 1349  
 Db 1344 ATACCTTGGCCCGCTGCTCCCAAGACGCTCTGTGGCGAGATCCGGTCCGCTGG 1403  
 QY 1350 AAGCATCTCTCTGTAGACGAAGCATGATTAAGCGCTTAAGGCCAATTCCTGGATC 1409  
 Db 1404 CAGCGACGACTCGCTGGCGAAGCGACGATGGCCAGCTTANGAGCCAGATCCGGGCTC 1463  
 QY 1410 GGGATGAGGTAAAGCAAGTGTAAAGCAGGACATGGGCTTCTGCTACTTTTANGAA 1469  
 Db 1464 GGGATGAGTGTCTACAGCTAGTTTTCGACCGCATGGCGGCGGTGCTGCTCGTGG 1523  
 QY 1470 CTCTGACAGCGCGGCGGTGCCAGCTGACGCTATAGCGTGCCTGACCAAAAAGTGC 1529  
 Db 1524 TAGCGACAAACGGCGCGGCGGCAACGGTGTGCTGCTCGCTGCGACGACATGCGGTG 1583  
 QY 1530 GGAGCTTAACAGCTCCA---GCATCTGCGAGGCTACTCMAACACTAGAAGTATCCA 1586  
 Db 1584 GGAGCTTAACAGCTCCA---GCATCTGCGAGGCTACTCMAACACTAGAAGTATCCA 1643  
 QY 1587 GGAGCTTAACAGCTCCA---GCATCTGCGAGGCTACTCMAACACTAGAAGTATCCA 1646  
 Db 1644 GGAGCTTAACAGCTCCA---GCATCTGCGAGGCTACTCMAACACTAGAAGTATCCA 1703  
 QY 1647 GCTGGCGGCTGTGGGTGTAGAAAAGCTGCAAGAGCTGCTGGCGATGAGGTGAGGT 1706  
 Db 1704 GCTGGGTGGCTGTGGCGCATAGAAAAGCAGCAAGCGGCTGGCGACACACGCT 1763  
 QY 1707 GCTTTTCAACCGGGAGCGAGTGCACCGCTGAGCAACCGATGTGGAAGCTTTTCA 1766  
 Db 1764 GCTTTTCAACCGGGAGCGAGTGCACCGCTGAGCAACCGATGTGGAAGCTTTTCA 1823  
 QY 1767 AGCACTAGAGCCAGCGCTGAGCGCTTTAGAACCTACATTAACCGAGCATTAAGATC 1826  
 Db 1824 GGTGCTGGAGCCCAAGGACAGATGGCTTCGGAACCTACTCGGAAAGGCAACCGCTTGC 1883  
 QY 1827 GCTGAGGAAATGCTGTAGACCGGGGACGCTTGTGCTCTTTGGGCAACGAAATGAC 1886  
 Db 1884 GCTGAGGAAATGCTGTAGACCGGGGACGCTTGTGCTCTTTGGGCAACGAAATGAC 1943  
 QY 1887 GCTTTTGGAGCGGCTGTGCTGACAGCGGCAACGCTTGTGCTGCTGAGTGGCTGAGTAC 1946  
 Db 1944 GGTGCTGCTGAGTGGCTGAGTGGCTGCTGCGGCAACCTAGCAAGCTTACGCTGGGCT 2003  
 QY 1947 GTTTTAAATAGCGGGTCACTATCCATGACTTCTTTGTAACCTGCTGAGCTCA 2006  
 Db 2004 GTTCACGAGGCTCTCGAGTGTACTGACACAGCACTTCTGCTGAGTGGCTGAGTGG 2063  
 QY 2007 CACTAATGCGAGCGCCGATGAATGACAGCAAGCTTTTGAAGAGAGCTTCAAAC 2066  
 Db 2064 TATCACTTGGAGCCCTCGGACAGAGATGAGGACCTACACGGGCAAGATGCG---AG 2120  
 QY 2067 TGCCCAAGTAAGTGGTGGCCACCGGCTAGACCTGATCTCGATCCAAATCCGAGCT 2126  
 Db 2121 TGCCCAAGTAAGTGGTGGCCACCGGCTAGACCTGATCTCGATCCAAATCCGAGCT 2180  
 QY 2127 AAGAGCCCTCGCAGAGGTGTACGGCTGTGCAGATCTCTGAAGAAAGTTTGTTAAGATT 2186

Db 2181 CCGGCGCTGTGAGAGCTTATGGCGGCTATGGCGGCTAGCCGACCGAGTTGCTGCAGCACTT 2240  
 QY 2187 TGTGAAGCGTGGCGCCCAAGTAATGAGCTGTGACCGGCTTGTATGTCG 2232  
 Db 2241 COTCGCTGCTGGGACAGGCTGATGAACCTCGACAGGCTTCGAGGTG 2286

RESULT 6  
 US-09-103-840A-2/C  
 ; Sequence 2, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: JOHNS, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; TITLE OF INVENTION: TUBERCULOSIS  
 ; FILE REFERENCE: 24366-20007-00  
 ; CURRENT APPLICATION NUMBER: US/09103,840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SEQ ID NO. 1: Patent In Ver. 2.1  
 ; SEQ ID NO. 2: 4403765  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; FEATURE:  
 ; OTHER INFORMATION: CDC 1551  
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 ; OTHER INFORMATION: represent a, t, c or g  
 US-09-103-840A-2

Query Match 37.7% Score 844.4; DB 3; Length 4403765;  
 Best Local Similarity 63.0% Pred. No. 3,2e-249;  
 Matches 1390; Conservative 0; Mismatches 801; Indels 15; Gaps 5;  
 QY 30 TACGTTATACACAACTACCGGAAATGCTTTTACCGGAGGTTGCTTTAAGCAAG 89  
 Db 2153346 TACAGAACACACACCGAGACCGCTAGACACGGCTGCTGCTGCTGGGTATATGAATTA 2153287  
 QY 90 TCGAGGTGGGGGACACAAAACAGGATGTTGGCCCAACATGCTCAACCTCGGCATCT 149  
 Db 2153286 TCGAGGTGGGGGACCGGAAACAGGATGTTGGCCCAACCGGCTCAATCTGAAGTACT 2153227  
 QY 150 AGCGGCTCATCTACGCTATGCGACCCCAACGACCGGATTTTGACTATCCGCAAGATT 209  
 Db 2153226 GCACCAAAACCGGCGTCTGACCGCTGCTGACCGCTGCTGCTGCTGCTGCTGCTGCT 2153167  
 QY 210 TAGAAGCTAGATCTGCGACGGGTTAAAGAGGCTTGGACGCTGATGACAGATCTCA 269  
 Db 2153166 GCGACATCATGAGTTGAGGCTTGGACGCTGATGACAGATCTCA 2153107  
 QY 270 GAGCTGTGGCCACGATTAAGGCTTATGTCGCCCTTCTTTATACCCATGGCTGCGCA 329  
 Db 270 GAGCTGTGGCCACGATTAAGGCTTATGTCGCCCTTCTTTATACCCATGGCTGCGCA 2153047  
 QY 330 CAGCGCGCGCCTACCTGATCGGTATGCGGCTGCTGGCGGCTGCTGCTGCTGCTGCTG 389  
 Db 2153046 CCGTGGCGGCTTACCGATCGCTGCGGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTG 2152987  
 QY 390 CTTCGCGGCTCTCAATCTGCGCAGCATGTCGCAATCTGATTAAGCAGCTGCTTCT 449  
 Db 2152986 GTTTCGCGGCTCTCAATCTGCGCAGCATGTCGCAATCTGATTAAGCAGCTGCTTCT 2152927  
 QY 450 TTGGGCCATCAACAAAAATACGGTGTGAAAAATCTCTCGCGGCTGCTTAATGACTAC 509  
 Db 2152926 TTGGGCCATCAACAAAAATACGGTGTGAAAAATCTCTCGCGGCTGCTTAATGACTAC 2152867  
 QY 510 AGGAAGCTAGCTCTGGAAGCTAGGCTGCTGGAAGCTGCTGGAAGCTGCTGGAAGCT 569  
 Db 2152866 CCGCAACTGCGGCTGGAAGCTAGGCTGCTGGAAGCTGCTGGAAGCTGCTGGAAGCT 2152807

QY 570 AGATGTATGGAGCTGGAAGAGATCTATCTGGGAGCAAGAACGATGGCGGAGA 629  
 Db 2152806 CGACAGTGGGAGGCC---CGATGAGGTCTATTGGGGCAAGGAACCACTGGCTCGCGGA 2152750  
 QY 630 CAGCGCTATGAAAGTGGACGAGAGCTCGAANAATCCCTGGAGCGGCTACAAATGGAGCT 689  
 Db 2152749 TGAAGCTATACAGCTGACAGGAGCTGTGAGNAGCCGCTGGCCGGTGGATGAGAGGCT 2152690  
 QY 690 CATTCTATTAACCCCGAAGCACCACAGCGACACCTGCTGCTGGTGGCGCTGCA 749  
 Db 2152689 GATCTACCTGATACCCGAGGGGCGCAAGGCAACCCGACCCCACTGGCGGGCGCTGA 2152630  
 QY 750 TATTCTGGAGACTTTGGCGGAATGGCAATGAATGACGAAGAAAGCGTGGCTCTCATAC 809  
 Db 2152629 CATTCCGAGAGTTTCTGGGCGCATGCGCATGAGAGACCTGGAACAGGGCGGCTGCTGT 2152570  
 QY 810 GGGTGGACACACTCTGGANAACCCATGCTGTGCTGATGGGAGAGATATCTGGGCGC 869  
 Db 2152569 CGGCGGTCAACACTTCTGGTAAAGACCAATGGGCG---CGGCGGGCGGATCTGGTGGCGCC 2152513  
 QY 870 AGAGCTTCGCGCGCGAGGTATTGAAGAAATGACGCTTGGGTTGGAAAACACCTACGGCAC 929  
 Db 2152512 GMAACCGAGAGTGGCTCTGCTGGAGATATGGGCTTGGAGGCTGGAGAGCTGTGGCAC 2152453  
 QY 930 CGACAGAGCTGGGATCATCAGCCATGAGCTAGAGAGGCGCTGGACCAAGACCCCTAC 989  
 Db 2152452 CGAACCGGCTTAAGGAGCGGATCACACAGCGGCATCGAGTGTATGGAGAACACCCGAC 2152393  
 QY 990 TCATGGAGCAATACCTTTTGGAAACCTCTTTGGTATAGAGTGGGAGCTTACCAAG 1049  
 Db 2152392 GAAATGGACAAAGTTCTCTGGATCTCTGAGGCTAGAGTGGAGCTGAGCAAG 2152333  
 QY 1050 TCCAGCTGGAGCTTATGATGGAACCAAAAGACGCTGCCGGGCTGGCACATACCGGA 1109  
 Db 2152332 CCCTGCTGGCGCTTGGCATATACCGCAAGAGCGCGCGCTGGCGCACATCCCGGA 2152273  
 QY 1110 TGCMAATGCCAGACATGCCAGCTCTCATTTATGCTCACTACGAGCTGGCGCTGGC 1169  
 Db 2152272 CCGGTGTC---GGGCGCGAGGGGCTCCCGAGGATCTTGGCCATGACTCTGCTGG 2152216  
 QY 1170 CATGGACCTGATTACGAANAAMATTTCTCGACGCTACTATGAACACCGCTGATGATTCG 1229  
 Db 2152215 GGTGGATCCGATCTATGAGCGGATCAAGGGTCTGGCTGGACACCCGAGGATATGGC 2152156  
 QY 1230 AGATGCTTTGGAAAGATGCTATGACACTGACACACAGATATGGACCAAGGTGCG 1289  
 Db 2152155 CGACGAGTTCGCAAGGCTTGTATGCTGATGCTCCAGGATAGGATGGGTCCGCTGGAG 2152096  
 QY 1290 CTACCTGGGACAGAGTGGCTCAGGAAGACCTCATCTGGCAAGACCTATACAGATGT 1349  
 Db 2152095 ATACCTTGGGCGGCTGTGCCCAAGCAGACCTGCTGTGGCAGGATCGGCTCGGGCT 2152036  
 QY 1350 AGGCACTTCTGTAGAGCAAGATATGAGGCTTAAGCGCTTAAGCAATTCCTGGATC 1409  
 Db 2152035 CAGCCACCTGCTGGCGAAGCGGATATGCCACCTATGAGCGAGATCTGGGAGT 2151976  
 QY 1410 GGGACTGACGAGTATCGACCTGTGAACGAGCATGGCTCTGCTCATCTTACTTTAGAAA 1469  
 Db 2151975 GGGATATGACTGTCTACAGCTATGTTTGGACGCAATGGCGGGCGGCTGCTGGTCCGGTG 2151916  
 QY 1470 CTCTGAGAGCGGGGCTGGCAAGGTGTGAGCTATGACATGCGGCCGCGCAAGAGCTG 1529  
 Db 2151915 TACGCAAGCGCGGCGCCGACGAGTGTGCAATGGCTTGGACGCAATGCTGGGTG 2151856  
 QY 1530 GGAAGTAAACACCTCA---GCACTTTCGAGGGTACTCAAAACACTAGAAGTATCCA 1586  
 Db 2151855 GGAAGTCAAGACCGCGAGGGGATCTGGCAGAGGTCAATTCGCACCTCGGAAGATCCA 2151796  
 QY 1587 GGAAGCTTTACACGGCGCATCATAGATACCAAGGAGTATGTTGGCGGACCTGATGT 1646  
 Db 2151795 GGAATCATTAATCCGCGGCGGGGAGATCATCAAGTGTCTTGGCGACCTCTGCTGT 2151736

QY 1647 GCTGGCGGCTGTGGGGTGTAGAAAAGTGCMAAAGATGCTGGCATGAGGTGAGGT 1706  
 Db 2151735 GCTCGGTGCTGTGGCGCCATAGAAAAGTGCMAAAGGCGCTGGCATGAGGTGAGGT 2151676  
 QY 1707 GCCTTTCAACCGGAGGAGGAGTGCACCGCTGAGCAACGATGCTGGGAAGCTTTGGA 1766  
 Db 2151675 GCGCTTACCGCGGGCGGACGATGGTGTGCGAAGAACAAACGAGCTGGAACTCTTGC 2151616  
 QY 1767 AGCACTAGACACAGGGCTGACGGCTTTAGAAACTATCAACCGAGGCAATAAAGTATC 1826  
 Db 2151615 GCTGCTGGAGCGAAGCGAGATGCTTCCGAACCTACTCGAAGAGGCAACCGCTGCC 2151556  
 QY 1827 CCCTGAGAGAATCTCTGTAGACGGGCGAGCTCTGTGCTGCTTGGCACCAAGAAAGAC 1886  
 Db 2151555 GCGCGAGTACATGCTGTGCAAGGGAACCTCTTACCTCAGTGGCTCCGCTGAGATGAC 2151496  
 QY 1887 TGCTTGGTATGGGATATGCTGTATGGGCAACCACTAGCAGGCTTGCAGACATGAGT 1946  
 Db 2151495 GGTGCTGGTGGCTTGGCGCTGTGCGGCAACTACGAGCTTACCGTGGGCTGT 2151436  
 QY 1947 GTTTTACAATAAGCGGGCTGTATCCAATGACTCTTTGTAAACCTCTGAGACTCAA 2006  
 Db 2151435 GTTCCACGGAGGCTCCGATCACTGACCAACGAGCTCTTGTTGAACTCTCGACATGG 2151376  
 QY 2007 CACTTAATGGCGAGGAGGATGATCAAGAACTTTTGAAGCGCAGAGCTTCAAAAC 2066  
 Db 2151375 TATCACTTGGAGCCCTCGCCACGAGATGACGGGACCTACCAGGCGAGATGAC 2151319  
 QY 2067 TGGCGAATTAAGTGGAGTGGCACCGCGGTAGACCTGATCTTGGGATCCAAATCCGAGCT 2126  
 Db 2151318 TGGCAGGTGAAGTGGACCGGCGAGGCTGGAGCTGCTTGGGCTCCCACTCGAGT 2151259  
 QY 2127 AAGAGCTTCCGAGAGCTTACGGCTGTGAGATCTTGAAGAACTTTGTTAAGATTT 2186  
 Db 2151258 GCGGGGCTTGTGAGCTCTATGCGCGATGAGCGGCAAGTTCGTGAGAGCTT 2151199  
 QY 2187 TGTGAAGGCTGGGCCAAAGTATGGACCTGGACCGGTTGTATGCTG 2232  
 Db 2151198 GCTGCTGCTGGGCAAGGTGATGACCTTCGACAGGTTGGAGCTG 2151153

RESULT 7  
 US-09-103-840A-1/c  
 ; Sequence 1, Application US/09103840A  
 ; Patent No. 6234328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; INVENTOR: FLEISCHMAN, Robert D.; FRASE, John R.; FRASE, John M.  
 ; APPLICANT: FRASE, John M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; TITLE OF INVENTION: TUBERCULOSIS  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103.840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; INVENTOR: FRASE, John M.; VENTER, John C.  
 ; SOFTWARE: PatentL Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4411529  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; OTHER INFORMATION: H37Rv  
 US-09-103-840A-1

Query Match 37.7%; Score 844.4; Db 3; Length 4411529;  
 Best Local Similarity 63.0%; Pred. No. 3, 2e-249;  
 Matches 1390; Conservative 0; Mismatches 801; Indels 15; Gaps 5;

QY 30 TACGCTATACACAAACACTGGCGGAAAATGCCCTTTTACGGAGGTTTGCCTTAAGCAAG 89  
 Db 2150806 TACAGAACACACCGGAGCGCTGAGCAACGGCTGCTCGGCTGTGGGTCATATGAATA 2156027  
 QY 90 TGCAGTGGCGGCAACAAACAGGATGTGGTGGCCCAACGCTCAACCTCGGCATCTT 149

Db 2156026 CCCCCTCGAGCGCGGGGAGACCTGGTGGCCCAACCGGCTCAATCTGAAGGTACT 2155967  
 QY 150 ACSCCAACATATCATGCTATTCGAGACCAACCGGAGTGTGAGCTTGCAGAGATT 209  
 Db 2155966 GCACCAAAACCCGGCGCTCCCTGACCCGATGGTGGCGGCTTCGACTATGCCCGGGGAT 2155907  
 QY 210 TAGAGCTAGATGTGCGACGGTTTAAAGAGACCTGGACGCTCAATGACAGATTAC 269  
 Db 2155906 CGGACACCTAGCTTGGTGGCCCTGACCGCGACGATGGAGGATGATGACACCTCGCA 2155847  
 QY 270 GGAAGTGTGGCAGAGATTACGCTCATATGCGGCTCTTTATAGCATGGCGTGGCA 329  
 Db 215846 GCGGTGTGGCGCGCGGACTACGCGCACTACGGGCGCTGTTATCGCGATGGCTGGCA 2155787  
 QY 330 CAGCGCGCGACCTACGCTATGTGTATGGCTGTGGCTGTGGCGGCTCCGCGCTCACGCG 389  
 Db 2155786 CCGTCTGGCGGCTCAATGCTACGCTACCGGCGCGCGCGCGCGGGGGGCGCATGAGCG 2155727  
 QY 390 CTTTGGCGGCTCTCAATGCTACGCTACCGGCGCGCGCGCGCGCGGGGGGCGCATGAGCG 449  
 Db 2155726 GTTGGCGGCTTTAAAGCTGTGGCGCGACCAACCGCACTGGATTAAGCAAGCTGTGCT 2155667  
 QY 450 TTGGCGGCTATGAGCAATATGCTGGCAAAATCTCTGGCGGATTAATGATATGAC 509  
 Db 2155666 GTGGCGGCTGAGAGATGACGCGCAAGACCTCTATGGCGGCACTATGTTTGGC 2155607  
 QY 510 AGGAAACGTAGCTTGAACATATGGGCTTTAAACATTTTGTGTTGACGCTGGCAGCG 569  
 Db 2155606 CGGCAACTGCGCTGGAATCGATGGCTTCAAGAGCTTCGGGTTCGGCTTCGGCGCGGT 2155547  
 QY 570 AGATGTATGGAGGCTCAGAGATGATATCTGGGCGAGAACCGAATGGCTGGGAGA 629  
 Db 2155446 CGACCACTGGGAGCG --- CGATGAGGTCTATTTGGGCGAAGGACCTCTGGCTGGC 2155490  
 QY 630 CAGCGCTATGAAGTGACCGAGAGCTCGAAATCCCTGGGAGCGCTGACAAATGGCAT 689  
 Db 2155489 TGACGCTTACAGCGGTATACGGGATCTGGAGACCGCGTGGCGCGGTGCGATGGGCT 2155430  
 QY 690 CATCTATGAACCGGAGCGACCGAGCGAGCGGAGAGCGCTATCGTGTCCGCTGA 749  
 Db 2155429 GATCTACGTGAACCGGAGCGCGAGCGGACCGGACCGCTATGGCGGCGGTGCA 2155370  
 QY 750 TATCTGTGAGACTTTGGCCGAATGGCAATGATGACGAGAAACCGGTGGCTCTCATGAC 809  
 Db 2155369 CATTTGCGAGCTTTGGGAGAGCGGCTGGGCAAGAGCTGCGTGAAGCGCGCTGATGCT 2155310  
 QY 810 GGTGCGACACCTTGGGAGAAACCGCATGCTGCGGATCGGAGCAATATGTGGCGCG 869  
 Db 2155309 CGCGGCTGACACTTTGGGTGAAGACCATGGCG --- CGCGCGCGCGCATTTGGGCG 2155253  
 QY 870 MAGCTGCGCGCGAGGTATTGAAGAAATGAGCTGGGCTGGAAACACCTACGCGAC 929  
 Db 2155252 CGAACCGAGCTCTCTCGCTGGAGAGATGGGCTGGGCTGGAAGCTCGTATGGCAC 2155193  
 QY 930 CGGACACGCTGGGATACATACACCGAGTATGAAGCGCGCTGACGACGACCGCTAC 989  
 Db 2155192 CGGACCGGCTTAGGAGGAGATACCGAGCGCATCGAGTGTATGACGACACCGCGAC 2155133  
 QY 990 TCATGAGGACATMACTTTGTGAAACCTCTTGTGGTACGAGTGGGAGCTTACCAAG 1049  
 Db 2155132 GAATGGGACACAGTTTCTCGAGATCTGTACGGCTGACGCTGAGGAGCTGAGGAAG 2155073  
 QY 1050 TCCAGCTGAGCTATTATCGTGGCAACCAAGAGCGCGGCGGTGGCCGACATCCGGA 1109  
 Db 2155072 CCGCTGTGGCGTTTGGCAATACACCGCCAGAGCGCGCGCTGCGGCACTCCGGA 2155013  
 QY 1110 TGACATATGACGAGCTGACGCTGCAATTTATGCTACTAGCGACCTGGCGGTGCG 1169  
 Db 2155012 CCGGTGCG --- GGGGGCGGGCGCTCCCGAGATGCTGGCACTGACCTCGGTGCG 2154956  
 QY 1170 CATGAGCCCTGATACGAAAAATTTCTCGAGGCTACTATGAACCCCTGATGATTTGCG 1229

RESULT 8

US-08-418-782-1

Db 2154955 GGTGTGCTCGATCTATGAGCGGATCAAGGCTCGCTGGCTGGAAACCCCGAGAAATGGC 2154896  
 QY 1230 AGATCTTTTGGCAAGCATGCTGTAAGATGACACAGAGATATGGGACCAAGAGTGGC 1289  
 Db 2154895 CGACGAGTGTGCGCAGGCGCTGGTACAGTATCCACCGACATGGGTCTTGTGGG 2154836  
 QY 1290 CTACTTGGGACAGAAAGTGGCTCTAGAGAGACCTCATCTGGCAAGACCTTACACGATGT 1349  
 Db 2154835 ATACTTGGGCGGTGGTGTCCCGACAGACCTCTGTGGCAGATCCGCTCCGCTGGCT 2154776  
 QY 1350 AGCCATCTCTGTGTAGACGCAACAGCATATTAAGGCTTAAAGCGCAAAATCTGGATC 1409  
 Db 2154776 CAGCCACGACCTGTGCGGAGCGGAGATTCGACGCTTAAGGCGCAGNATCCGGCATC 2154716  
 QY 1410 GGGATGAGGTAAGCGATGGTGGTAAAGCAGGATGGGCTTCTGCATCTACTTTTAAAGAA 1469  
 Db 2154715 GGGATGAGGTAAGCGATGGTGGTAAAGCAGGATGGGCGGGGCTGTGCTTCGCTGG 2154656  
 QY 1470 CTCGTGACAGCGGCGGCTGACCTGACCTGACGCTCAGCATATGATGGCCGACCAAGAGCTG 1529  
 Db 2154655 TAGGCGACAGCGGCGGCGGCGCAACGGTGTGCGATCTGCTGACGCGACCAAGCTGGT 154596  
 QY 1530 GGGATTAACCAAGCTCA --- GCATCTTGGCAGGCTACTCAAAACACATAGAGATATCCA 1586  
 Db 2154595 GGAGTCAACGCCCGGAGGAGATCTCCGAGGTTCTATCGCAAGCTTATCGCACCTGGAAGATCCA 2154536  
 QY 1587 GGAGGACTTTAAACAGGCGCACTAGATTAACAAAGCATGTCTGTGGCGACCTGATGCT 1646  
 Db 2154535 GGGATCATTAACCTCCGCGGCGCGGAGCAATCAAGGTCTCTTCGCGGACCTGTGCT 2154476  
 QY 1647 GCTGGCGGCTGTGGGCTGTGAAGAAAGCTGCAAAAGTCTGGCCATGAGGTGAGCT 1706  
 Db 2154475 GCTGGTGGCTGTGGCCCTAGAGAAAGACACGAGGCGCTGGCCACATCATCGGT 2154416  
 QY 1707 GCTTTCAACCGGCGAGGAGTCCACCGCTGAGCAACCGATGTCGAGCTTTCGA 1766  
 Db 2154415 GCGCTTCAACCGGCGGCGGCGGAGTGGTGGCAGCAACCAACGACCTGGAATCTTTCG 2154355  
 QY 1767 AGCATCTAGGCGGCTGACGGCTTTAGAACTACATTAACCGGAGCATAAAGTATC 1826  
 Db 2154355 CGTCTGGAGCCCAAGGACATGGCTTCCGAAATCTCTCGAAGGCGCACCCCTTGGC 2154296  
 QY 1827 CCGTGAGGAATGCTCGTATAGACCGGCGGCGAGCTTCTGCTTTTCGGCACCAAGAAATGAC 1886  
 Db 2154295 GCGCGAGTACTGCTGTGCAAGAGGACATGCTTACGCTCAGTGGCCCTCGAGATGAC 2154236  
 QY 1887 GTCTTTGTGAGGCGCTGCTGCTAGTGGGACCAACTAGCAGCGTTTGCAGGATGAGT 1946  
 Db 2154235 GTGTCTGTGATGGCTGCGCTCGCGCAAACTACAGAGCTTACCGTGGCGCT 2154176  
 QY 1947 GTTTAGAAATAGCGGCTGAGTATCCATGACTCTTTGTAACTGCTGAGCTCAAA 2006  
 Db 2154175 GTTACCGGGCTCTCGAGTCACTGACCAAGCACTTCTGTGAACTGCTGCATATGGG 2154116  
 QY 2007 CACTAAATGGCCAGCAGCGATGATCAGCAAAAGTTTTGAAGCGAGAGCTTCAAAAC 2066  
 Db 2154115 TATCACTTGGAGCGCTTGGCAGCATAGCAGGACCTACAGCGCGGAGGTGGC --- AG 2154059  
 QY 2067 TGGGAGTAAAGTGGAGTGGACCCCGGTGAGACTGATCTTCGATCAATTCGAGCT 2126  
 Db 2154059 TGGCAAGTAAAGTGGACCGGCGCGGTGGACCTGGTCTTCGGCTCAATTCGAGTT 2153999  
 QY 2127 AAGAGGCTTCCAGAGATGTAGCGCTGTGCGATCTGTGAGAAAGTTTGTTAAGATT 2186  
 Db 2153998 GCGGGCGCTTGTGAGGCTTATGGCGCGCATGCGGACCGGCAATTTCTCGAGCTT 2153939  
 QY 2187 TGGGAGGCGCTGGCCCAAGTAAATGACCTGGACCGGTTTGAATCTG 2232  
 Db 2153938 CCGCTCTGCTGGGACAGGATGAAGCTCGACAGGTTTCGACGTG 2153893

[illegible]

Query Match 37.2%; Score 832.4; DB 1;  
Best Local Similarity 63.0%; Pred. NO. 2.7e-247;  
Matches 1389; Conservative 0; Mismatches 801;  
Indels 16; Length 2235;

Query Match	Best Local	Score	DB 1	Length
Matches 1389; Conservative	Similarity 63.0%; Pident 34.7%; Mismatches 801; Indels 16; Gaps 6			
OY	30	TACGTATATACAAACACATGGCGAAATATGCCCTTTATTCACGAGAGTTCGGTAAACGAAAG	89	
OY	30	TACGTATATACAAACACATGGCGAAATATGCCCTTTATTCACGAGAGTTCGGTAAACGAAAG	89	
OY	90	TGCGAGCGCGCACACCAAAACAGAGATTCGGTCCCAACAGTCGCTCACTTCGGATCT	149	
DB	93	CCCGCTCGAGGCGCGGGAACACAGAGATTCGTGGCCCAACCGCTCAATCTGAAGTACT	152	
OY	150	AGCCGACATATCATGCTNTTCGGACCCAAACGACCCGGATTTTGACTATGCGAAGATTT	209	
DB	153	GCACCAACCCGCGCTTCGCTGACCCGATGGGTGGCGGTTCGACTATGCGGAGGT	212	
OY	210	TAAAGAACTAGATCTGGCAGCGGTTTAAAGAGACCTCGACGCGCTTAATGACAGATTCACA	269	
DB	213	CGGACCATCTGAAGCTTGACGCCCTCTGACGGGACCATCGAGAGAGTGATGACCATCTGCA	272	
OY	270	GGACTGTGGCCGCGGACATACGCTGACTATTPGCCCTTCTTTTATACGATGGGTGGCA	339	
DB	273	GCGCTGTGGCCGCGGACATACGCTGACTATTPGCCCTTCTTTTATACGATGGGTGGCA	339	
OY	330	CAGCGCGGACACTACGCTATGGGTGATGGCTGGTGGCGGTGGCTCGGGCTCAACAGC	382	
DB	333	CGCTCGGCGGACCTACGCTATCCACAGCGCGCGCGCGCGCGCGCGCGCGGCGGATCGACGC	392	
OY	390	CTTTGGCGGCTCTCAATAGCTGGCCGACAGATGGCATCTGCTGAATAGACAGCTTGCTTCT	449	
DB	393	GTTTGGCGCGGCTTTAAGACTGGCCGACACAGCGACAGCTTGGACAGGCGCGCGGCTGCT	452	
OY	450	TTGGCGCCATCAACAAAAATACGGTTCGAAAAATCTCTCGGCGGATCTTAATGATACCTAC	509	
DB	453	CGACAGCTCTGACAGAACTGATGGGACAGAGCTCTCATGGGCGGACCTGATGTTTTCGC	512	
OY	510	AGGAACTGCTCTGAGAACTATGGGCTTTAAAGCTTTCTGGTTTCTGGTTCGAGTGGCAGAC	569	
DB	513	CGGCAACCTCCGCGCTGGAACTCGATGGGCTTCAAGAGCTTCGCGTTCGGCCGGGT	572	
OY	570	AGATGATGGGAGCTCAAGAAGATGTATATCTGGGAGCAGCAAAACCAATGGCTGGGAGA	629	
DB	573	CGACGATGGGAGCC---CGATAGGTCTATTGGGCGAGAGAGCAACCTGCTTCGGGCA	629	
OY	630	CAAGCGCTTGAAGCTGACCGAGAGCTCGAAATCCCTCGGGACCGCTGCAAAATGGGACT	689	
DB	630	TGAAGCTTTACAGCGGTAAAGCGGATCTGGAGAACCCGCTTCGCCGGGTGCAGATGGGCT	689	
OY	690	CATCTATGTAAACCCCGCAGGACCCAGCGCAAGCCAGACCTCTATCGCTGCTGGCGGTGA	749	
DB	690	GATCTATGTAAACCCCGCAGGACCCAGCGCAAGCCAGACCTCTATCGCTGCTGGCGGTGA	749	
OY	750	TATTCGTGTAGACTTTTGGCGGCTAGCAATGAATGACAGAAACCCGCTCTCTACAG	809	
DB	750	CATTTCGCGAGCTTTTCGCGGCTATGGCCATGAAGAGAGTCTGAACAACGCGGCGCTGATGCT	809	





Db	33	TACAGAACACACCGAGCGCCCTAGCAGCGGCTCTCCGCTGGGTCAATATGAMTA	92	Db	1107	CCCGTTGC---GGGGCCAGGGCGCTCCCGAGCATGCTGGCCACTGCTCTCGCTGGG	1163
Qy	90	TCGAGCTGGGGGACCAAAACAGGGATGTGGGCCCAACATGCTCAACCTCGGCATCTT	149	Qy	1170	CATGGACCTGATATGCAAAAATTTTCTGAGGGTACTATGAACCTCTGATGATTTGG	1229
Db	91	CCCGCTCGAGGGCGCGAAACAGGACTGGTGGGCCAAGCCGGCTCAATCTGAAGTACT	152	Db	1164	GGTGGATCCGATCTATGCGGGATCAGCGTCTGCGGACACCCCGAGGATATGG	1223
Qy	150	AGCGCACTATCTGCTATCGGACCAACAGCCCGGATTTTGATCTGCGGAGAGTT	209	Qy	1230	AGATGCTTTTCGGAAACGCTGTACAACATGACACAGATATGGGACCAAGCTGGG	1289
Db	151	CGGCGGCGGCGGCTGCTGACCGGATGGGTGGCGCTTCGATCTGCGCGGAGGT	212	Db	1224	CGAGAGTTTGGCAAGGCTTGGTACAGCTGATCCACAGAGCATGGGTTCGCTTGGGAG	1283
Qy	210	TACAGAGCTAGATCTGCGACGGTGTAAAAGAGCTCGGACAGCGCTTAATGACAGATTCA	269	Qy	1290	CTACTCTGGGACGAGTCCCTCTAGGAGAGCTCATCTTGGCGAGACCTCTATACAGATGT	1349
Db	213	CGCGACCATACAGCTTGGGCCCTCGAGCGGAGACATCGAGGAAGTATGACCACTCGCA	272	Db	1284	ATACCTTGGCGCGCTGGTCCCAACGACCTGCTGTGTGGAGATCGCTCTCGGCT	1343
Qy	270	GGAGCTGGGCGGACAGATATAGGCTATATGAGGCGCTTAAGGCGCTTAAGGCGCTTAAGG	329	Qy	1350	AAGCCATCTCTTGTAGACGAAAGAGATTTGAAGGCTCTAAAGCGCTTAAGGCGCTTAAGG	1409
Db	273	CGCTTGGGCGGCTTACAGCTTGGCCGACACAGCGGCTTGTATCGGATGGGCTTGGG	332	Db	1344	CAGCAGACCTCTGCTGGGAGGAGAGATTTGCGGCTTAMAGAGAGATCCGGGCTC	1403
Qy	330	CAGCGCGCGGACCTACCGCTATCGGTATGCGGCTGGTGGGCGGCTCTCCGCTCAACAGG	389	Qy	1410	GGGATGACGCTTAAAGCGCTGTAAAGCGGCTTAAGCGGCTTGAAGCTTCTTGAAT	1469
Db	333	CGTCTGGGCGGCTTACAGCTTGGCCGACAGAGCGGCTTGGGCGGCGGCGGCGGCGG	392	Db	1404	GGGATGACCTTGTCTACAGCTAGTTTTCGACGATGCGGCGGCGGCTGCTGCTTGGG	1463
Qy	390	CTCTGGGCGGCTTCAATAGCTTGGCCAGCAATGCTTGAATGAGCAAGCTTGGCTTCT	449	Qy	1470	CTCTGACAGCGCGGCTGGCAAGCGGTGACGCTATAGCATGCGCCCAACAAAAGACTG	1529
Db	393	GTCTGGCGCGCTTACAGCTTGGCCGACACAGCGCTTGTGACAGAGGCGGCGGCTGCT	452	Db	1464	TAGCAGACAGCGCGGCGGCGGCGGAGCTGCTGCGCTTGGCGCTTGGCGGCTGCTGCT	1523
Qy	450	TTGGCCCATCAACAAAATAGCGTGGAAATCTCTCTGGGCGGATCTAATGATATCTAC	509	Qy	1530	GGAGTTAAACACCTCTCA---GCNACTTGGCAGGCTACTCAANACTATAGAGATTTCA	1586
Db	453	GTGGCGGCTCAAGAGAGATACCGCAAGAGCTCTGATGGCGGCGGCTGATTTTTCG	512	Db	1524	GGAGTTCAAGGACCCGCGGAGATCTGGCGGAGCTCATTCGACCTGGAAGAGATCCA	1583
Qy	510	AGGAACTAGCTCTGAAATCTTGGCTTAAACTTTTGGCTTTTTCGAGCTGGCGAGCT	569	Qy	1587	GGAGCTTTTACCGAGCGCATCAGATACAAAGAGTATCTGTTGGCGGACTCTGATGT	1646
Db	513	CGGCACTTGGCGGCTTGAATGATGATGGCTTCAAGAGCTTCTGGGTTGGGCTTGGCGG	572	Db	1584	GGATCTTACCGCGGCGGCGGAGCTGCTGCGAGACACCGGCTGAGTCTTCTTGG	1643
Qy	570	AGATGTATGGAGGCTGAAGAGATGTATCTGGGAGCAGAAACCGAATGGCTGGGAGA	629	Qy	1647	GCTGGCGGCTGGGCTGTAGNAAAGCTCAAAAGATGCTGGCCATGAGGCTCGAGT	1706
Db	573	CGACAGTGGGAGCT---CAGTGGGCTTATGGGGCAGGAGCCGCTGCTGGCTCGGCA	639	Db	1644	GCTGGCTGGCTTGGCGGCTATAGAGAAAGCAGCAAGCGGCTGGCCACACATCCG	1703
Qy	630	CNAGCGCTATGAGGTGACGAGAGCTCGAANAATCCCTGGGAGCCGCTACAAATGGGACT	689	Qy	1707	GCCTTTTAAACCGGCGAGGGATGCGCACCGCTGAGCAACCGATGGAGCTTTTGA	1766
Db	630	TGAGCGTTTACGCGGCTTACGCGGCTTGGAGAACCCGCTGGCGCGGCTGCAGATGGGCT	689	Db	1704	GGCTTTTACCGCGGCGGCGAGATGCTGCGAGACACACCGGCTGAGTCTTCTTGG	1763
Qy	690	CATCTATGTAACCCGAGAGACCCACCGCAGCCAGCTATGCTGCTGGCGCTCA	749	Qy	1767	AGCACTAGAGCCAGCGCTGACCGCTTAAAGAACTACATTAACCGGAGCATTAAGTATC	1826
Db	690	GATCTAGTACAGGAGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGCGGCTGCA	749	Db	1764	GCTGCTTGGAGCCCAAGGAGATGGGTTTCGAAACTACCTCGGAAGGCAACCCCTTGGC	1823
Qy	750	TATGCTGTAGCATTTTGGCGAATGGCAATATGACAGAGAAACCGTGGCTCTCATAGC	809	Qy	1827	CGTGAAGATGCTCTGTAAGCGGCGGCGCATCTTCTGCTTTTGGGACCAAGATGAC	1886
Db	750	CATTTGGGAGAGTTTGGGCGGTGCGCATGAACAAGCTCGAAACAGCGGCGCTGATCT	809	Db	1824	GGCGGCTACNTGCTCTGCAACGGGCGATGCTTACGCTAGTGGCTCGCTGATGAC	1883
Qy	810	GGTGGAGACACTCTGGGAAACCCATGCTGTGCGCATGCGGAGAAATATGGGCGG	869	Qy	1887	TGCTTTGTAGAGCGGCTGTGTGATGATGGGACCACTAGCAGCTTGGCAGCATGGAT	1946
Db	810	CGGCGCTCACCTTCTTGTAGNAGAACGGCG---CGCGCGCGCCCATCTGTGTCGG	866	Db	1884	GGTGTCTGTATAGTGGGCTGGGCGCTCTCGG---GCNAACTTCAAGCGGTTACCGTGGCG	1942
Qy	870	AGAGCTCGGCGCGCATTTTGAAGAAATGAGCTGGGCTGGGCTGGAAACCACTACG	929	Qy	1947	GTTTTAAATAAGCGGCTCAGCTATCCAAATGACTCTTTGTAAACCTGCTAGACCTCA	2006
Db	867	CGAACCTCGAGGCTCTCGGCTGGAGCATGGGCTTGGCTGGAGAGCTCGTATGACAC	926	Db	1943	GTTCACGAGGCTCTCGATCTGCTGACCAAGACTCTCTGTGATGCTGCTGATCTG	2002
Qy	930	CGGACAGCGTGGGATACATACCATAGTGGACTAGAGGCGCTGGACCAAGCCCTTAC	989	Qy	2007	CATTAATGGCCAGCGCATGATCAGCAAGTTTGTGAGCGGAGGAGCTTCAAAAC	2066
Db	927	CGGACCGGTNAGGCGGATACACAGCGGCTGAGGTGCTTGGAGCAACCCCGAC	986	Db	2057	TATCAGCTTGGAGGCGCTGGCGAGCATGACGGGACCTACAGGCAAGGATGGC---AG	2059
Qy	990	TCAATGAGCATATCTTTTGAACCTTTTGGTACGAGTGGGAGGTTTACCAAG	1049	Qy	2067	TGGGAGTAAATGGTGGAGTGGCAACCGGGTAGACTGATCTTCGGATCAATTCGAGCT	2126
Db	987	GAATGGGACACAGATTTCTCGAGATCTCTGACGGCTAGAGTGGGAGTCAAGAG	1046	Db	2060	TGGAGGTTGGTGGGCGGCGGCGGCTGGACTGGTCTTTCGGGCTCACTCGGTT	2119
Qy	1050	TCCAGCTGGAGTTTATCATGTGGAACCAAAAGACGGTGGCGGCTTGGCACCACTCGGA	1109	Qy	2127	AAGAGCCTCGCAGAGAGTACGCTGTCAGATCTTGAAGAAAGCTTCTTAAGATT	2186
Db	1047	CTCTCTGGGCTTGGCATCTACCGCGAGGAGCGGCGGCTGGCGGCACTCCCGG	1106	Db	2120	CGGGCGCTTGTTCGAGGCTATGGCGGCAATGACGCGAGCGGAGTTCGTGCGAGCTT	2179
Qy	1110	CGCATGATATCCCAAGTGGCAGCTCATTTATGCTCATACGGAACCTGGGCTGCTG	1169	Qy	2187	TGTGAGGCTTGGGCGCAAGTATGAGCTGAGCGGCTTTGATCTG	2232
Db				Db	2180	CGTGGCTGGCGAGAGTGTATGCTTGGAGGCTTGGAGCTG	2225

RESULT 11  
 US-08-418-782-2  
 : Sequence 2, Application US/08418782  
 : Patent No. 5658733  
 : GENERAL INFORMATION:  
 : APPLICANT: Cockerill, Franklin R.  
 : APPLICANT: Cockerill, Franklin R.  
 : APPLICANT: Cockerill, Franklin R.  
 : APPLICANT: Cockerill, Franklin R.  
 : TITLE OF INVENTION: Detection of Isoniazid Resistant Strains  
 : TITLE OF INVENTION: Detection of Isoniazid Resistant Strains  
 : NUMBER OF SEQUENCES: 22  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Schwesman, Lundberg & Woessner  
 : ADDRESS: 3500 IDS Center  
 : CITY: Minneapolis  
 : STATE: MN  
 : COUNTRY: USA  
 : ZIP: 55402  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : SOFTWARE: PC-DOS/MS-DOS  
 : CURRENT APPLICATION DATA:  
 : APPLICATION DATA: Sequence #1.0, Version #1.25  
 : FILING DATE: US/08418,782  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Woessner, Warren D.  
 : ADDRESS: 1501 Hennepin Ave., Suite 1000  
 : REFERENCE/DOCKET NUMBER: 150.141U51  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 612-339-0331  
 : TELEFAX: 612-339-0361  
 : INFORMATION FOR SEQ ID NO: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : TYPE: nucleic acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: DNA  
 : US-08-418-782-2

Query Match 32.38; Score 721.8; DB 1; Length 2221;  
 Seq. Loc. Similarity 60.54; Pred. No. 4.9e-213;  
 Matches 1335; Conservative 0; Mismatches 842; Indels 28; Gaps 8;

390 CTTGGCCCTCTCAATAGCTGCCAGCAATGCAATCTGGATAAAGACGCTTGCTCT 449  
 393 GTTTCGGCCGCTTACAGCTGCGCGACACCCAGCTTGGACAGGCGCGCGCTC 452  
 450 TTGGCCCATCAACAAACAAATACGCTCGAATAATCTCTGGCGGATCTAATGATCACTGC 509  
 453 GTGGCCCGGTCAAGAGAAGTACGCGAAGAGCTCTCATGGCGGACTGATTTGTTTTCGC 512  
 510 AGAAGAC--GTACTCTGGAACATATGGGCTTTTAAACTTTTGGTTTTCAGTGGCAGA 567  
 513 CGGCATCTCCCTGGCTGGGATTCGATGGGCTTCAGACGCTTCGGGTTCGGCTCGGGCT 572  
 568 GCAGATCTATGGAGGCTCGAAGAAGATGTAATCTGGGAGCAGCAACCAAGTCTGGGGA 627  
 573 CGACCATGGGAGAC-----GATGAGGTCTATTGGGCGAAGAGGACCTTGGCTCGGC 627  
 628 GATGA--CGGTACAGCTGANGCNATCGAGACCCCTGGCGGGGTTCAGATGGG 684  
 688 CTCATCTATGTAAACCCCGGAGACCAACGGCAGCAGACCTCTATGCTGCTGGCGGT 747  
 685 CTGATCTAGTGAACCGGAGCGCGGACGCAACCCGACCCCATATGCGCGCGGGCTC 744  
 748 GATATCTGGAGACTTTTGGCGATGAGCAATGATGAGGAAGAACCGTGGCTCTCAT 807  
 745 GACATTCGGAGAGCTTTGGCGGCAATGGCCATGACGCTCGACAGCGGCGCTGATC 804  
 808 GCGGCTGGACACCTTCGGAACAAACCCATGGTGTGCGGATCGGAGAAATATGTGGGC 867  
 805 GTGCGCGGTCACTTTCTGGTAAGACCCATGG---CGCGGCGCGCGCGGCTGCTGGCGC 861  
 866 GAGAGCCTTCGCGCGCAGGATATGAGGAATGAGAGCTGGGTTGGAACCACTTACGCG 927  
 862 CCGAACCAGGCTGCTCGCTCGAGCAATGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCT 921  
 928 AC CGGACGACGCTGCGGATACCATCAGTGAAGTCTGAGAGGCGCTGAGCAAGACCCCT 987  
 922 AC CGGACGCGTGAAGAGCGGATCACAGCGGATCGAGGCTGATGAGAGCAACCGCG 981  
 988 ACTCAATGAGCAATCACTTTTGAAGACGCTTGTGTTACGAGTGGAGCTTACGAA 1047  
 982 ACGAATGGGACACAGTTTCTCTGAGATCTGTACGGCTACGAGTGGAGCTGACGAAG 1041  
 1048 AGTCCAGCTTGAAGCTTATCTAGTGAACAAACAGCGTGGCGGGCTGGCACCATACCG 1107  
 1042 AGCGCTTGGCTGAGTATGACACCGCAAGCGCGCGGCTGGCGGACCATACCG 1101  
 1108 GATGCGACATGATCCGAGCTGCGGATGAGTCTGATGATGATGATGATGATGATGATGATG 1167  
 1102 GACCGGTTGC--GGGGGCGCGGCGCTCCCGGAGTCTCGGCACTGACCTGCTGCTGCT 1158  
 1168 GCGATGAGCGCTTATGAGAAAAATTTCTGACGCTTACTATGAACCCCTGATGAGTTT 1227  
 1159 GCGGTTGCTGATCTATGAGGGGCTACGCGTCTGGCTGGACACCCCGAGGAATTG 1218  
 1228 GCGAGTCTTTCGCGAAGCTGTACAACTGACAGCAACGAGTGGAGTGGAGTGGAGTGG 1287  
 1219 GCGAGAGTTCGCGAAGCTGTACAACTGACAGCAACGAGTGGAGTGGAGTGGAGTGG 1278  
 1288 GCGTACCTGGGACAGAGTGGCTTACGAGAGCTCATCTGGCAAGACCTTATACAGAT 1347  
 1279 AGATACCTTGGGCGGCTGGTCCCGACAGACCTCTGCTGGCAGATCGCGTCCGCTCGC 1338  
 1348 GTGAGCCTCTCTTGTGAGCGAAGCAATGATGAGCTTAAAGCGCAATTCGGA 1407  
 1339 GTCAGC---ACGACCTCTGCGGACAGCAGATTCGACGCTTAAGGCGCAGATCCGSCA 1395  
 1408 TCGGAGCTGAGGCTGAGGAGCTGTGAGCAGCGCATCGGGCTTCTGCACTACTTTTGA 1467  
 1396 TCGGAGTACCTGCTTCACAGATGTTTTCGACCGCATGCGGCGGCGGCTGCTGCTGCTG 1455  
 1468 ACCTTCACAGCGCGGCGGCTGCCACGCTGCTGACGCTTATGACTGGCGCCACAAAGAC 1527



QY 748 GATATCTGCGAGCTTTTGGCGGAGTGGGATGATGAGCAAGAAACCGTGGCTCTCTATA 807  
 Db 749 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 810  
 QY 750 GACATTCGCGAGAGCTTTTGGCGGCATGGCCATGAGCAGATCGAAGAACACCGCGCGTATC 804  
 Db 808 GGGGGTGAGCAACACCTTTGGGAAAAACCATTTGGTCTGCGCATGCGGAGAAATATGTGGGC 867  
 QY 809 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 811  
 Db 810 GTCCGGGCTGCACACTTTTGGGTAGAACCATATG---CGCGCGCGCGCGCATCTTGGTGGCG 861  
 QY 868 CGAGAGAGTGGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 921  
 Db 869 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 922  
 QY 870 CCGAACCCGAGGCTGCTCCGCTGGAGAGATGGCTTGGCTTGGGATGAGATGAGATGAGAT 927  
 Db 928 ACCGAGACGGTGGGATACCATACAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 981  
 QY 929 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 990  
 Db 930 ACCGAGACGGTGGGATACCATACAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 981  
 QY 988 AGCTACATGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1047  
 Db 989 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1048  
 QY 990 AGCAAAATGGGACACACTTTTCTCGAGATCTGTACGGCTAGAGTGGGAGCTGACGAG 1041  
 Db 1048 AGTCCAGCTGGAGCTTATCAGTGGGAAACCAAAAGAGCGTCTCGGGCTGGCACCAATACG 1107  
 QY 1049 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1110  
 Db 1050 AGCTCTTGGGCTGGGATACCATACAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1101  
 QY 1102 GATCCATGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1167  
 Db 1103 GATCCATGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1168  
 QY 1104 GATCCGCTTGC---CGGGGACAGGGGCGCTCCGAGCATGCTGCGCATGACCTCTCGCTG 1158  
 Db 1168 CGCATGGACCTGATATACGAAAAAATTTCTCGAGGCTACTATGMAAACCTGATGATTT 1227  
 QY 1169 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1230  
 Db 1170 CGGATGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1218  
 QY 1228 GCAGATGCTTTTGGGAAAGCATGTCAGAACTGACACAGAGATATGGGACCAAGGTG 1287  
 Db 1229 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1288  
 QY 1230 GCAGATGCTTTTGGGAAAGCATGTCAGAACTGACACAGAGATATGGGACCAAGGTG 1287  
 Db 1231 GCAGATGCTTTTGGGAAAGCATGTCAGAACTGACACAGAGATATGGGACCAAGGTG 1278  
 QY 1288 CGCTACTGGACCAAGATGGCTTCAGGAAGACCTCATCTGGCAAGACCTATACAGAT 1347  
 Db 1289 AGTACTGTTGGGCTGGTCCGAGAGACCTCTGCTGGAGAGATCCGGTCTCGCTGGC 1338  
 QY 1348 GTAGGCACTCTTGTGAGCAAAACAGTATTTGAAGCGCTTAAGAGCAAAATCTCGGAA 1407  
 Db 1349 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1408  
 QY 1350 GTACG---ACGACCTCTGCGGGGCAAGCAGATATGCCACCTTANAGAGCAGATCGGGCA 1395  
 Db 1408 TGGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1467  
 QY 1396 TGGGATGATGCTTCTACACTGTTTGGCCGATATGGGGGGGGGGTGGTCTGCTG 1455  
 Db 1456 AACTCTGACAGCGGGGCTGCACAGTGGCAGGTATAGATAGCTGGCCCAAAAGAC 1527  
 QY 1456 GTGAGCAGACAGCGGGGCGGCGCAAGCTGGTGGCATCGCTCGACAGCACAGTGGG 1515  
 Db 1528 TGGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1587  
 QY 1587 TGGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1575  
 Db 1588 GAGAGCTTTTACAGCGGCAATATGATCAAGCAAGTATGCTGGCGAGCTGATGCTG 1647  
 QY 1576 GAGTATCA-----CTGGCGCGGGAACATCAAAAGTGTCTTTGGCGGCTGCTGCTG 1629  
 Db 1648 CTGGCGGCTGCTGGGCTGAGAAAGCTGCAAAAGATGCTGGCATGAGGTGAGGTG 1707  
 QY 1630 CTGGGCTGCTGCTGGGCTGAGAAAGCTGCAAAAGATGCTGGCATGAGGTGAGGTG 1699  
 Db 1708 CTTTCAACCGGAGCGGATGCGCATCGCTGAGCAACAGTGTGGAGCTTTGGAA 1767  
 QY 1690 CCGTTCACCGCGGGCGCGAGATGCGGTGCGAGCAAAACCGAGTGGATCTCTTGGC 1749  
 Db 1768 GCACTAGAGCCAGCGCTGACGGCTTTAGAAACTACATTAACCGGAGCATAAAGTATCC 1827  
 QY 1750 GTGCTGGAGCCGAGCAGATGGCTTCCGAACTACCTCTCGGAAGGGGACCGCTGGCGG 1809

QY 1828 GCTGAGAAATGCTCTGAGACCGGGCGAGCTTCTGCTTCTTGTAAACCTGCTGACCTCAAC 1887  
 Db 1810 CCGAGTACCTGCTGCTGAGCGGAGCACTGCTTACGCTCAGTGCCTGAGATGACG 1869  
 QY 1888 GCTTTGGTAGCGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1947  
 Db 1870 GTGCTGGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1925  
 QY 1948 TTTTAAATTAAGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2007  
 Db 1930 TTTTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1989  
 QY 2008 ACTAATTAAGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2067  
 Db 1990 ATCACTTGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2046  
 QY 2068 GCGGAGTAAGTGAAGTGGCAAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2127  
 Db 2047 GCGGAGTAAGTGAAGTGGCAAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2106  
 QY 2128 AGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2187  
 Db 2107 CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2166  
 QY 2188 GTGAAGGCTTGGCGCAAGTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2232  
 Db 2167 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2211

## RESULT 13

US-08-952-219-2  
 ; Sequence 2, Application US/08852219  
 ; Patent No. 592575  
 ; INVENTOR: COCKERILL, FRANKLIN R.  
 ; APPLICANT: COCKERILL, FRANKLIN R.  
 ; APPLICANT: KLINE, BRUCE C.  
 ; APPLICANT: UHL, JAMES R.  
 ; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains  
 ; TITLE OF INVENTION: of M. Tuberculosis  
 ; NUMBER OF SEQUENCES: 22  
 ; CONFIDENCE ADDRESS: 22  
 ; ADDRESS: 22  
 ; STREET: 119 No. 592575th Fourth Street, Ste. 203  
 ; CITY: Minneapolis  
 ; STATE: MN  
 ; COUNTRY: USA  
 ; ZIP: 55401  
 ; COMPUTER READABLE FORM:  
 ; COMPUTER FILE: 41k  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/952, 219  
 ; FILING DATE: 07-May-1997  
 ; CLASSIFICATION: 435  
 ; NAME: SANDHURST, VICTORIA A.  
 ; REGISTRATION NUMBER: 41,287  
 ; REFERENCE/DOCKET NUMBER: 230.00010130  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 612-305-1228  
 ; TELEFAX: 612-305-1228  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SOURCE: CHAIRMAN'S OFFICE  
 ; LENGTH: 2211 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA  
 US-08-952-219-2

Query Match

32.3%; Score 721.8; DB 2; Length 2221;

Best Local Similarity 60.5%; pred. No. 4.8e-213;  
Matches 1335; Conservative 0; Mismatches 842; Indels 28; Gaps 8;

[illegible]







Db 2422 GTGGCCGCTGACGAAGAGTACGCGCAGCAAGCTCTCATGGCGCGACATGATGTTTTCGC 2481  
 QY 510 AGGAAC--GTAGCTCTGGAACATATGGCTTTAAACCTTTGTTTTGTCAGGTGCGCA 567  
 Db 2482 CCGCAACCGCTGCGCTCGCATCGATGGCGTTTACAGAGTTCGGGTTCGGCTTCGGCGT 2541  
 QY 568 GCAGATGTATGGAGCTCGAAGAATGATATCTAGGGAGCAGAAACGAAATGGCTGGGA 627  
 Db 2542 CGACAGTGGGAGCC-----GATGGGTCTATTTGGGGCAGGAGGCCCTGGCTCGGC 2596  
 QY 628 GACAAAGCCTATGATGAGTACCGAGAGCTCGAAATCCCTCGGGGCGGTACAAATGGA 687  
 Db 2597 GATCA---CGGTATAGCTATAGCGATCTGAGAACCGCTGGCGCCGCTCCAGATGGGG 2653  
 Db 688 CTCATCTATGTAAACCGAAGACCAACGCGCAGACCTATCGCTGCTCGCGT 747  
 QY 2654 CTCATCTAGTACCGGAGCGGCGGCGACACCGCGGCGCCGATCGCGGCGGCTGTC 2713  
 Db 748 GAAATCTAGTACGATCTTTGGCGGATGCAATGATGACGAGAAAGACGTGGCTCTCAT 807  
 QY 2714 GACATTCGCGAGAGCTTCGCGCGCATGGCGCATGACGACCTCGAAAGCGGGCGTGCAT 873  
 Db 808 GGGGTGGACACACCTTCGGAAMACCATGTGCTCGCGATGGCGGAGAAATATGTGGG 867  
 Db 2774 GTGGCGGTCACTTTTCGTAGACATCATG--CGCGGCGCGCCGCTCTGTGTGGC 2830  
 QY 868 CGAGAGCTTCGCGCGGATGATGAGAAATGAGCTTGGGTGAGGAGAAACCTACGGC 927  
 Db 2831 CCAGAACCGAGAGGTCTCGCTGGAGCAGATGGGCTTGGGCTGGAAAGCTCGTATGGC 2890  
 QY 928 ACCGACAGGTGCGGATACCATCATACGTGGACTAGAGGGCGCTTGGACCAAGACCC 987  
 Db 2891 ACCGACAGGTGCGGATACCATCATACGTGGAGCGCATCGAGTCTGATGGAGACACCGC 2950  
 QY 988 ACTCAATGGAGCACTAACTTTTGGAAACCTCTTTGGTACGAGTGGGAGCTTACCAA 1047  
 Db 2951 ACGAAATGGGACACAGTTTCTCGAGATCTGTACGGCTACGAGTGGAGCTGACAG 3010  
 QY 1048 AGTCAGCTGGAGCTTACTGGAACACAGAGAGGTGCGGGCTGGACCATACGC 1107  
 Db 3011 AGCCCTGCTGGCGCTTGGCATACACGCGCAGGACGCGCGGTCGCGCAGCATCCG 3070  
 QY 1108 GATGCATATGTCAGCAAGTCCGACCGCTCCATTTATGCTCACTACGCACTGGCGCTG 1167  
 Db 3071 GACCGGTGCG---GGCGGACAGGGCGCTCCGACGATGCTGGCCACTGACCTTCGCGT 3127  
 QY 1168 GCGATCGCTGATACGAAATTTCTCGAGCTACTATGAAACCTGATGAGTTT 1227  
 Db 3128 CGGGTGGATCGGATCATACCGGATACCGCTGCTGGCTGGAGCAACCCCGAGGAGTT 3187  
 QY 1228 GCAGATGCTTGTAGACATAGTACAACTGTACACACAGAGATATGGCAAAAGTG 1287  
 Db 3188 GCGCAGAGTTCGCGAAGGCTGTGACAGCTGATCCACGAGACATGGGTCGCGTGGC 3247  
 QY 1288 CQCTACCTGGGACAGATGCTGCGAGAGCTCATCTGCGAAGAGCCCTATACAGAT 1347  
 Db 3248 AGATACCTTGGCGCTGCTCCCAAGACGCTTCTGTGGCAGGATCCGCTCCCTGGC 3307  
 QY 1348 GTAAGCATCTCTTGTAGAGCAAAACGATATTGAAGGCTTAAGAGCCAAATACTGGAA 1407  
 Db 3308 GTCAGC--AGACATCTGGCGGAGAGGATATCCAGCCCTTAAGAGCCGATCGCGGCA 3364  
 QY 1408 TGCGGACTGAGTACGAGCTGTAGCGGAGTGGCTGTGATCTACTTTAGA 1467  
 Db 3365 TCGGGATGTGCTCATAGCTAGTTTGTAGCCGATAGTGGCGGCGGCTGCTGTTCCGT 3424  
 QY 1468 AACTCTGACAGCGCGCGCTGCAAGCGGTGACAGCTATAGAGCTGGCCCAAAAAGAC 1527  
 Db 3425 GGTAGCAGAGCGGCGGCGGCGAGGTGTCGATTCGGCTGCGAGCCACACATCGGG 3484  
 QY 1528 TCGGAGTAAACACGCTACGACCTTGCAGGCTACTCAAAACATAGAGGTATCAG 1587  
 Db 3485 TGGGAGTCAACGACCCGACGGATCTGCGCAGGTCTATTCGACCTCGAGAGATCCAG 3544

QY 1588 GAGGACTTTAACGAGCGGCAATCATGATACGATAGTATGCTTGCGCCGACCTGATGTC 1647  
 Db 3545 GAGTCTATCA-----CTGCGCGCGGAGATATCAAGTGTCTTGTGCGGACCTCTGTC 1698  
 QY 1648 CTGCGCGGCTGTGCGGCTGTAGAAAAGCTTGCAAAAGATGCTGGCCATATGAGTGCAGTG 1707  
 Db 3599 CTGCGTGGCTGTGCGGCTATAGAAAGACGAAAGAGGCGCTGCGCAACATCATCGGTG 3658  
 QY 1708 CTTTTCATCTGGGACGACCGGATGCAACCGCTGACCAACGATGTGGAAGCTTTGGA 1767  
 Db 3659 CCCTTCACCGCGGCGCGCAGATGTGGTGGCAGGCAACACGAGCTGGGATCTTTC 3718  
 QY 1768 GCATCTGAGCGGCTATGGGCTTTAGAAAATCATATTAACCGGAGCATAAAGTATCC 3827  
 Db 3719 GTGCTGGAGCCGCAAGGCAGATGGCTTCGGAACCTNCTCGGAAGGGCAACCGTTGCCG 3778  
 QY 1828 GCTGAGGAATGCGGTAGACCGGGGAGCTTCTGCTTTGCGACCGACGAAATGACT 1887  
 Db 3779 CGGAGTACATCGCTGCTCGAAGCGGACGCTGCTTACGCTGCTGCTGAGATGACG 3838  
 QY 1888 GCTTTGCTGAGCGGTATGGGTGACTGGGCACCACTACGACGCTTGGCAGCATGGAGTG 1947  
 Db 3839 GTGCTGCTGAGTGGCTTGGCGGCTCTCGGCGCAACTACAGCGCTTACCGCTGGGCGTG 3898  
 QY 1948 TTTACAAAGGCGGGCTCACTATACGATGAGCTCTTTGTAAGCTGCTAGACTCMAC 2007  
 Db 3899 TTCACGAGGCTCGGAGTCACTGACCAACGACTTCTCTGAGCTGCTGACATGGGT 3958  
 QY 2008 ACTAAATGGGAGCGCAGCATGAATCAGCAAAAGTTTTTGAAGCGAGAGCTTCAAACT 2067  
 Db 3959 ATCACTGGGAGGCTCTGGCCAGCATGACGGGACCTACCGGCGAAGGATGGC---AGT 4015  
 QY 2068 GCGCAGTAAAGTGTGTCGCAACCGCGGTGAGACCTGATCTGAGTCGATTCGAGCTA 2127  
 Db 4016 GSCAAGTGAAGTGGACCGCGACCGCGTGGACCTGCTTGGCTCAACTCGAGTTG 4075  
 QY 2128 AGACCCCTCGAAGAGTGCAGGCTGTGCAGATTTCTGAAGAAAAGTTTGTAAAGATTT 2187  
 Db 4076 CGGGCGCTTGTGAGGCTATGTGCGCGATGACGCGGACGGAGTTTGTGACAGGATTC 4135  
 QY 2188 GTGAAGGCTGGGCCAAAGCTAAAGGACTTGAAGCTGGACCGGTTTGAAGTCTG 2232  
 Db 4136 GTCGCTGGCTGGCAACAGTGTATGAACCTCGACAGGTTGCGAGGTG 4180

Search completed: October 8, 2003, 15:34:05  
 Job time : 137.323 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: October 8, 2003, 12:53:54, Search time 572.431 Seconds  
(Without alignments)  
1044.788 Million cell updates/sec

Title: US-09-884-889-7

Perfect score: 2238

Sequence: 1 atggaataacacaaactc.....accgtttgatctgaataa 2238

Scoring table: IDENTITY\_NWC

Gapop 10.0, Gapext 1.0..

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubna/us07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubna/us06\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubna/us05\_PUBCOMB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubna/us06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubna/us07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubna/PT05\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubna/PT08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubna/PT08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubna/US07A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubna/US07A\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2238	100.0	2238	10	US-09-884-889-7
c	2238	37.7	2993	15	US-10-080-170-645
3	844.4	37.7	2993	15	US-10-080-170-645
4	844.4	37.7	2993	15	US-10-080-170-645
5	251.8	11.3	620	11	US-09-940-925A-145
c	251.8	11.3	620	11	US-09-940-925A-145
6	251.8	11.3	620	11	US-09-940-925A-145
7	250.2	11.2	620	11	US-09-940-925A-143
c	250.2	11.2	620	11	US-09-940-925A-143
9	250.2	11.2	620	11	US-09-940-925A-146
c	250.2	11.2	620	11	US-09-940-925A-146
11	250.2	11.2	620	11	US-09-940-925A-150
c	250.2	11.2	620	11	US-09-940-925A-150
13	250.2	11.2	620	11	US-09-941-193A-143
c	250.2	11.2	620	11	US-09-941-193A-143
14	250.2	11.2	620	11	US-09-941-193A-146
c	250.2	11.2	620	11	US-09-941-193A-146
16	248.6	11.1	620	11	US-09-940-925A-144

c	17	248.6	11.1	620	11	US-09-940-925A-148	Sequence 148, App
c	18	248.6	11.1	620	11	US-09-941-193A-144	Sequence 144, App
c	19	248.6	11.1	620	11	US-09-941-193A-148	Sequence 148, App
20	157.2	7.0	391	10	US-09-825-574-1	Sequence 1, App11	
21	157.2	7.0	391	12	US-09-882-945A-1	Sequence 3, App11	
22	155.6	7.0	391	10	US-09-825-574-3	Sequence 4, App11	
23	155.6	7.0	391	10	US-09-882-945A-4	Sequence 4, App11	
24	155.6	7.0	391	12	US-09-882-945A-3	Sequence 4, App11	
25	155.6	7.0	391	12	US-09-882-945A-2	Sequence 2, App11	
26	154.6	6.9	391	12	US-09-825-574-2	Sequence 2, App11	
27	154	6.9	391	12	US-09-882-945A-2	Sequence 2, App11	
28	38.8	1.7	4301	10	US-09-917-800A-1666	Sequence 1666, App	
c	30	37.4	1.7	9025508	14	US-10-205-194-136	Sequence 126, App
31	36.8	1.6	1049	12	US-10-156-761-1	Sequence 1, App11	
32	36.8	1.6	1049	12	US-10-140-472-358	Sequence 358, App	
33	36.8	1.6	1049	12	US-10-141-761-358	Sequence 358, App	
34	36.8	1.6	1049	12	US-10-142-885-358	Sequence 358, App	
35	36.8	1.6	1049	14	US-10-123-155-358	Sequence 358, App	
36	36.8	1.6	1049	15	US-10-146-731-358	Sequence 358, App	
c	37	36.8	1.6	10470	8	US-08-781-986A-20	Sequence 20, App1
c	38	36	1.6	520	14	US-10-184-644-332	Sequence 332, App
c	39	36	1.6	520	14	US-10-184-644-332	Sequence 332, App
c	40	36	1.6	3523	14	US-10-108-877-1	Sequence 1, App11
41	35.8	1.6	548	13	US-10-027-632-166215	Sequence 166215, App	
42	35.8	1.6	548	13	US-10-027-632-166215	Sequence 344, App	
43	35.8	1.6	888	12	US-10-141-761-544	Sequence 544, App	
44	35.8	1.6	888	12	US-10-142-885-544	Sequence 544, App	
45	35.8	1.6	888	12	US-10-158-790-544	Sequence 544, App	

#### ALIGNMENTS

#### RESULT 1

US-09-884-889-7  
; Sequence 7, Application US/09884889  
; Patent No. US20020102680A1  
; GENERAL INFORMATION:  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: DIVERSA CORPORATION  
; APPLICANT: SANVAL, Indrajit  
; APPLICANT: ADHIKARI, Robert  
; TITLE OF INVENTION: CATALASES  
; FILE REFERENCE: DIVER100-4  
; CURRENT APPLICATION NUMBER: US/09/884,889  
; CURRENT FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US/09/412,347  
; PRIOR FILING DATE: 1997-10-16  
; PRIOR APPLICATION NUMBER: US/08/951,844  
; PRIOR FILING DATE: 1997-10-16  
; PRIOR APPLICATION NUMBER: US/08/674,887  
; PRIOR FILING DATE: 1996-07-03  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; SEQ ID NO 7  
; TYPE: DNA  
; ORGANISM: Microscilla furvescens  
US-09-884-889-7

Query Match 100.0%; Score 2238; DB 10; Length 2238;  
Best Local Similarity 100.0%; Pred. No. 0;  
.Matches 2238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 ATCGAAATACAAACACTCAGGATCTTCTACGATATACCAACACTCGCGGAAATGCG 60  
Db 1 ATCGAAATACAAACACTCAGGATCTTCTACGATATACCAACACTCGCGGAAATGCG 60  
Oy 61 CCTTTTACCGAGGTTGCTTAAAGCAAAATGCGGCGGACCAAAAACAGGGATTGG 120  
Db 61 CCTTTTACCGAGGTTGCTTAAAGCAAAATGCGGCGGACCAAAAACAGGGATTGG 120

QY 121 TGGCCACATGCTGACACCTGGGCACTTATGCGCAACATTTATCGCTATCGGACCCAAAC 180  
 Db 121 TGGCCACATATCTCACTCTGGGCACTTATGCGCAACATTTATCGCTATCGGACCCAAAC 180  
 QY 181 GACCCCGGAATTTTGAAGCTATGCCGAAGCTTTAAGAGCTAGATCTGCGAGCGTTTAAAGAG 240  
 Db 181 GACCCCGGAATTTTGAAGCTATGCCGAAGCTTTAAGAGCTTTAAGAGCTTTAAGAG 240  
 QY 241 GACCTGGCAGCGCTATGACAGATATACAGAGCTTGTGGCAGCAGATTAAGCTCATAT 300  
 Db 241 GACCTGGCAGCGCTATGACAGATATACAGAGCTTGTGGCAGCAGATTAAGCTCATAT 300  
 QY 301 GCCGCCCTCTTTATACGATAGGCTGGCAGAGCGGCGCACTACGATATCGGTATGGC 360  
 Db 301 GCCGCCCTCTTTATACGATAGGCTGGCAGAGCGGCGCACTACGATATCGGTATGGC 360  
 QY 361 GCGTGTGTCTTTATACGATAGGCTGGCAGAGCGGCGCACTACGATATCGGTATGGC 420  
 Db 361 GCGTGTGTCTTTATACGATAGGCTGGCAGAGCGGCGCACTACGATATCGGTATGGC 420  
 QY 421 GCCAACTTGTGAATAAGCAGCGTGTCTTTTGGCCCATCAACAAAAATACGCTCGAA 480  
 Db 421 GCCAACTTGTGAATAAGCAGCGTGTCTTTTGGCCCATCAACAAAAATACGCTCGAA 480  
 QY 481 ATCTCTGCGCGCATCTATGATATGATATGATGATGATGATGATGATGATGATGATG 540  
 Db 481 ATCTCTGCGCGCATCTATGATATGATATGATGATGATGATGATGATGATGATGATG 540  
 QY 541 AAAAATTTTGGTGTGGAGTGGCAGAGCAGATGATGGGAGCTTGAGAGAGATGTATAC 600  
 Db 541 AAAAATTTTGGTGTGGAGTGGCAGAGCAGATGATGGGAGCTTGAGAGAGATGTATAC 600  
 QY 601 TGGGAGCAGAACCGATGCTGGGAGCAGAGCAGATGATGGGAGCTTGAGAGAGATGTATAC 660  
 Db 601 TGGGAGCAGAACCGATGCTGGGAGCAGAGCAGATGATGGGAGCTTGAGAGAGATGTATAC 660  
 QY 661 ATCTCCCTGGGAGCGCTACAAATGGGAGCTCATCTATGTAACCCCGAGGACCCAGCG 720  
 Db 661 ATCTCCCTGGGAGCGCTACAAATGGGAGCTCATCTATGTAACCCCGAGGACCCAGCG 720  
 QY 721 AAGCAGACCCCTATGCTGCTGCGGTGATATGCTGAGACATTTGGCGCGAATGGCAATG 780  
 Db 721 AAGCAGACCCCTATGCTGCTGCGGTGATATGCTGAGACATTTGGCGCGAATGGCAATG 780  
 QY 781 ATGAGCAGAACCGCTGCTCTCATATGCGGGTGGACACATCTTGGGAAAAACCCATGGT 840  
 Db 781 ATGAGCAGAACCGCTGCTCTCATATGCGGGTGGACACATCTTGGGAAAAACCCATGGT 840  
 QY 841 GCTGGCGATCGGAGAAATATGTTGGCGAGAGCTGACCGCGCGCATATTGAGAAATG 900  
 Db 841 GCTGGCGATCGGAGAAATATGTTGGCGAGAGCTGACCGCGCGCATATTGAGAAATG 900  
 QY 901 AGCTGGGGTGGAAAAACATACGCGACCGGACAGGTGGGATACCATACGATGGA 960  
 Db 901 AGCTGGGGTGGAAAAACATACGCGACCGGACAGGTGGGATACCATACGATGGA 960  
 QY 961 CTAGAAGGCCCTGGACACAGACCCCTACTCAATGGAGCAATTAATTTTTGAAACCTC 1020  
 Db 961 CTAGAAGGCCCTGGACACAGACCCCTACTCAATGGAGCAATTAATTTTTGAAACCTC 1020  
 QY 1021 TTTGGTATCAGGTGGAGCTTACAAAGTTCAGTGGAGCTTATCAGTGGAAACCCAAA 1080  
 Db 1021 TTTGGTATCAGGTGGAGCTTACAAAGTTCAGTGGAGCTTATCAGTGGAAACCCAAA 1080  
 QY 1081 GAGGTGCGGGGGTGGACCATACCGATGACATGATGCGGACGATGCGACGATGCGCA 1140  
 Db 1081 GAGGTGCGGGGGTGGACCATACCGATGACATGATGCGGACGATGCGACGATGCGCA 1140  
 QY 1141 TTTATGCTCACTACGAGCTGGCGCTGCGGATGGAGCTGATATACGAAAAATTTTCGA 1200  
 Db 1141 TTTATGCTCACTACGAGCTGGCGCTGCGGATGGAGCTGATATACGAAAAATTTTCGA 1200  
 QY 1201 CGGTACTATGAAACCCCTGATGAGTTTTCGAGATGCTTTGGCGAAGCATGTGACAACTG 1260

Db 1201 CGGTACTATGAAACCCCTGATGAGTTTTCGAGATGCTTTGCGGAAGCATGTGTAACAACTG 1260  
 QY 1261 ACACACAGATATGGGACACCAAACTGCGCGTACCGTGGGACAGATGTGGCTTGAAGAAC 1320  
 Db 1261 ACACACAGATATGGGACACCAAACTGCGCGTACCGTGGGACAGATGTGGCTTGAAGAAC 1320  
 QY 1321 CTATCTTGGCAGACCTATACCAAGATGATAGCCATCTCTTGTAGAGAAAAACGATAT 1380  
 Db 1321 CTATCTTGGCAGACCTATACCAAGATGATAGCCATCTCTTGTGTAGAGAAAAACGATAT 1380  
 QY 1381 GAGGCGCTAAAAGCAAAATCTCGGATCGGAGCTGAGCGTAAAGCGTGGTAAACAG 1440  
 Db 1381 GAGGCGCTAAAAGCAAAATCTCGGATCGGAGCTGAGCGTAAAGCGTGGTAAACAG 1440  
 QY 1441 GCATGGGCTCTGATCTACTTTTAAAGAACTGTGACAAAGCGCGCGGTGCCAACGGTGA 1500  
 Db 1441 GCATGGGCTCTGATCTACTTTTAAAGAACTGTGACAAAGCGCGGTGCCAACGGTGA 1500  
 QY 1501 GGTATACGATGGGCGCCACAAAGAGCTGGGAGTAAACAACTTCAGCACTTGCACGG 1560  
 Db 1501 GGTATACGATGGGCGCCACAAAGAGCTGGGAGTAAACAACTTCAGCACTTGCACGG 1560  
 QY 1561 GTACTCAAACTAGTAGAGGTATCCAGAGAGACTTTAACCGGCGGCAATCAGATACAA 1620  
 Db 1561 GTACTCAAACTAGTAGAGGTATCCAGAGAGACTTTAACCGGCGGCAATCAGATACAA 1620  
 QY 1621 GCATGATCTGGCGACCTGATTTGGCTGGCGGCTGCGGGGTGTAAGAAAGCTGCA 1680  
 Db 1621 GCATGATCTGGCGACCTGATTTGGCTGGCGGCTGCGGGGTGTAAGAAAGCTGCA 1680  
 QY 1681 AAGATGCTGGCATGAGGTGAGGTGCTGCTTCAACCGGAGCAGCGGATGCCACGCT 1740  
 Db 1681 AAGATGCTGGCATGAGGTGAGGTGCTTCAACCGGAGCAGCGGATGCCACGCT 1740  
 QY 1741 GAGCAACCGATGAGAGCTTTTGAAGCACTAGAGCCAGCGCTGAGCGCTTGAAGAAC 1800  
 Db 1741 GAGCAACCGATGAGAGCTTTTGAAGCACTAGAGCCAGCGCTGAGCGCTTGAAGAAC 1800  
 QY 1801 TACATTAACCGGAGCATTAAGTATGCGCTGAGAAATGCTGCTAGACGGGCGAGCTT 1860  
 Db 1801 TACATTAACCGGAGCATTAAGTATGCGCTGAGAAATGCTGCTAGACGGGCGAGCTT 1860  
 QY 1861 CTGTGCTTTTGGCACCAAGAAATGACTGCTTGTAGCGGCTGATGGTGTACTGGGCA 1920  
 Db 1861 CTGTGCTTTTGGCACCAAGAAATGACTGCTTGTAGCGGCTGATGGTGTACTGGGCA 1920  
 QY 1921 TACTAGCAGCTTGGCAGCATGCTGCTTACAAATAGCGCGCTGATGCTATGCTATGAC 1980  
 Db 1921 TACTAGCAGCTTGGCAGCATGCTGCTTACAAATAGCGCGCTGATGCTATGCTATGAC 1980  
 QY 1981 TTTCTTTGTAACCTGCTAGACTCAACACTTAATGGCGACCGCAGCGATGATGAGCAAA 2040  
 Db 1981 TTTCTTTGTAACCTGCTAGACTCAACACTTAATGGCGACCGCAGCGATGATGAGCAAA 2040  
 QY 2041 GTTCTTTGTAACCTGCTAGACTCAACACTTAATGGCGACCGCAGCGATGATGAGCA 2100  
 Db 2041 GTTCTTTGTAACCTGCTAGACTCAACACTTAATGGCGACCGCAGCGATGATGAGCA 2100  
 QY 2101 CTGATCTGGGATCCAAATTCGAGCTAAGAGCTTGAAGGCTTGGCGGCTGAGAGTGTAG 2160  
 Db 2101 CTGATCTGGGATCCAAATTCGAGCTAAGAGCTTGAAGGCTTGGCGGCTGAGAGTGTAG 2160  
 QY 2161 TCTGAGAAAGCTTCTTAAGCTTTTGTGAGGCTGGGCGCGCAAGTATGCGGACTGGAC 2220  
 Db 2161 TCTGAGAAAGCTTCTTAAGCTTTTGTGAGGCTGGGCGCGCAAGTATGCGGACTGGAC 2220  
 QY 2221 CGGTTTGTATGAAATAA 2238  
 Db 2221 CGGTTTGTATGAAATAA 2238

RESULT 2

QY	750	FATTTCGCGAGCATTTGGCGAATACATGATCTACTGCAAGAAACCGTGCCTCTCAGC	809
Db	39758	CATTTCGCGAGAGCTTTGGCGGCATGGCCATGAACGAGCTCGCAACACGCCGCTGTATCST	39659
QY	810	GSGTGGAACAACCTTCGGAATAAACCCATGTGCTGGCATATCCGAGAATATGTGGGGCG	869
Db	39698	CGGGCTGTCANCTTTTGGTAGACATCCATGGGCG---GGCGCCGGCGGCATCTGGTCGGCG	39642
QY	870	AGACGCTGCGCGCCGATGATGAAATGATGAGATGAGATGAGATGAGATGAGATGAG	929
Db	39581	CGGAGCTGTAAGACGCGATACACGCGGCATAGAGGTGCTATGGACMACCCCCGAC	39522
QY	990	TCAATGAGACATAACTTTTTTGAANAACCTCTTGTTGATGCTACAGCTGGAGCTTTCACAAA	1049
Db	39521	GAATGGGACACACATCTTCTCTGAGATCTCTGTACGCTACGAGTGGGAGCTGACGAGAG	39462
QY	1050	TCCAGCTGGAGCTTATCAGTGGAAACAAAAAGACGGTCGGGGGCTGGCACCATACCAGA	1109
Db	39461	CCCTGCTGGCTTGGCATACACCCACAGAGCGGGCGCTGGTGGCMACCCCAGGANTTGGC	39402
QY	1110	TGCATGCCATCCCAAGCTGCGACAGCTCCATCTGCTGCTACGTACGACATGGGCGCTCG	1169
Db	39401	CCCGTTGCG---CGGGCGAGGGCGCTCCCGCCAGCATCTTGGCCCATGACCTCTCGCTGCG	39345
QY	1170	CATGACCTCGATTACGAAAAAATTTCTCGAGGCTACTATGAACACCTCATGAGTTTGC	1229
Db	39344	GGTAGATTCGATCTATGACGGGTACAGCTGCTGCTGGTGGCMACCCCAGGANTTGGC	39285
QY	1230	AGATGTTTTCGGAAGACATGTTACAAACTACACACAGATATGAGGACCAAAGTGGC	1289
Db	39284	CGAGAGAGTTCCGCAACAGCCTGGTACAAAGCTGATCCACGAGACATGGTCCGCTTGGCAG	39225
QY	1290	CTACTGGGACAGAGCTCTCCGAGAGACTCATCTGGCNAGACCTTATACCAGATGT	1349
Db	39224	ATACTTGGGCGCTGTTCCCAACAGACCTTCCGAGGATATCGGCTCGATGGT	39165
QY	1350	AGACATCATCTTGTAGACGAAACAGATATTAGAGGCTTAAAGCAAAATCTCGTAATC	1409
Db	39164	CAGCGACAGATCTCTGCGGGAAGCCGAGATTCGCAGCTTAAAGGCCNATCGGGGCATC	39105
QY	1410	GGGACTGAGGTGACGAGCTGTGPAAGACGAGATGGGCTTGTGCATCTACTTTAGAA	1469
Db	39104	GGGATGACTGCTCAGCAAGTAGTTTCGACGGCATGGGCGGCGGTGCTCTCGTGG	39045
QY	1470	CTCTGACAAGCGGCGGGTGCACAGGTGCAGCTATACGATGGCCGCCCAAAAGACTG	1529
Db	39044	TGAGACAGAGGGCGGGCGGCGAGGTGGCATCCGCTGCGACGACNATCGGGGT	39895
QY	1530	GAGATTAACACACCTCCA---CGACCTTCGAGGTACTCAAAACCTAGAGATATCCA	1586
Db	38984	GGAGTCAACGACCCCGGAGGEGATCTCGCAGCGTCATCGCANCCCTGGAGAGNTCCA	38925
QY	1587	GGAGGACTTAAACGAGCGCAATCAGATAACAAAGCATGTCTGTGGCGCACTGATGT	1646
Db	38924	GGGATCAATCAATCCGCGGGGCGCGGGACATCAAAAGTGTCTCTCGGCGCATCTGT	38865
QY	1647	GTCGCGGCTGTGGGTGTGTAAGAAAGTCTGCAANNAGTGTGGGCGCATGAGTGCAAGT	1706
Db	38864	GCTCGGTGGCTGTGCGGCATAGAGAAGACCAAGCGGCTGGCCACACATCATCGAT	38805
QY	1707	GCCTTTCAACCCCGGAGAGCGGATGCCACCGCTGACCAAAACGATGTGGAACTTTGA	1766
Db	38804	GGCCTTCACCCCGGGCGCAGGATGGCTGCAGAGCAAAACGAGTGGATCTCTTTC	38745
QY	1767	AGCTACTGACAGCGGCTGAGCGGTTTGAMCTTACTATTAAACGGGACATTAAGTATC	1826
Db	38744	GGTCTGGAGCGATGSCCTTCCGAAATCTCTCTGCGTAAGGGGCAACATCTGAC	38685
QY	1827	CSGTGAGGAAATGCTGTGATAGCGGGCGSAGCTGTCGTGGTAAAGGCACTAAATGAC	1886

```

Db      38684  GCCGAGTACATGCTGCTGCACAGCGGCAAGCTCTTACGCTAGTGGCCCTGAGTGC 38625
QY      1887  TCGTTTGTAGCGGTATACGCTGTACTGGCGCAACCACTACGACGTTGCGCAGCATGAGT 1946
Db      38624  GGTCTGTGTAGTGGCTGCGCGCTGCTCGGCGCAACTACAGAGCGCTTACCGCTGGGCGT 38565
QY      1947  GTTTCANATATAGCGCGGTAGCTATCAATGACTTCCTTTGTAACCTGCTAGACTCAA 2006
38564  GTTCCAGGCGCTCGCATCATGACACAGCTTATGACACAGCTTCTTGGACGCTGCAATGG 38505
QY      2007  CACTAAATAGCGGACGAGCATGATCAGACAGAAATTTTGTGAGCGGAGACTTCCAAAC 2066
Db      38504  TATCACTGGGAGCCCTCCGCCAGAGATACAGGGGACAGATGCG---AG 38448
QY      2067  TGGCGACTTAATAGTGGATGGCCACCGGCTAGACCTGATCTTCGATATCATTCGAGCT 2126
Db      38447  TGGCAAGTGAATGGACGGCAGCGCTGCACTGCTCTCGGCTGCACTCGGAGTT 38388
QY      2127  AAGAGCCCTCGCAGACTTACGGCTGTGAGATCTTCGAAGAAATTTGTTAAGAGTTT 2186
Db      38387  GCGGGGCGCTGTGTCGAGGCTATATGGCGCGGATACGGCAGCGGAAGTGTGTCAGGACT 38328
QY      2187  TCGGAGGCGCTGCGGCAAGTATGAGCTATGAGCTGGACGGCTTTGATCTG 2232
Db      38327  CTTGCTGCTCGTGGAGAGGTGATGCAAGCTCGACGTTCCAGCTG 38282

```

## RESULT 3

US-09-884-889-5

: Application US/09884889

: Patent No. 5, US/002010280A1

: GENERAL INFORMATION:

: APPLICANT: DIVERSA CORPORATION

: APPLICANT: ROBERTSON, Dan

: APPLICANT: SANTAL, Indrajit

: APPLICANT: ADHIKARI, Robert

: FILER: INVENTION CATALASES

: FILER REFERENCE NUMBER: US/09/884,889

: CURRENT APPLICATION NUMBER: US/09/884,889

: PRIOR FILING DATE: 2001-06-19

: PRIOR APPLICATION NUMBER: US/09/412,347

: PRIOR FILING DATE: 1999-10-05

: PRIOR APPLICATION NUMBER: US/09/951,844

: PRIOR FILING DATE: 1997-10-16

: PRIOR FILING DATE: 1997-08/674,887

: PRIOR FILING DATE: 1996-07-03

: NUMBER OF SEQ ID NOS: 8

: SOFTWARE: PatentIn version 3.0

: SEQ ID NO 5

: LENGTH: 2262

: TYPE: DNA

: ORGANISM: Alcalligenes (Deleya) aquamarinus

US-09-884-889-5

Query Match 30.0%; Score 672.2; DB 10; Length 2262;

Best Local Similarity 59.3%; Pred. No. 3.2e-209;

Matches 1277; Conservative

; Mismatches 808; Indels 60; Gaps 6;

QY 99 CGCGACCCAAAAGAGGATGGTGGCCCAACTGCTCAAGCTCGGCTATTAAGCAACA 158

Db 150 CGGATCTCCCAAGATATGGTGGCGAGGTTGACGTTATTTGATCAACGCA 209

QY 159 TTGATCGTCACTCGGACCAACGCGGATTTGATATGCGGAGAGTTTAAGAGCT 218

Db 210 AGATGCGCAATCAGACCGGATGATTCGGATTTCACTACCGTCAAGAAGTACGAGCT 269

QY 219 AGATCTGCGACGGTTAAAGAGACCTGGCAGGCTTAATGACAGATTCACAGGACTGGT 278

Db 270 CAAATTTCAGCGCTGACGAAGATGTCACGGCTGTGATGACCGATAGCCAGATGGTG 339

QY 275 CCGACAGATACGTCGCTTATATGCGCTCTTTATACGATGCGCTGCGCAGCGCG 338

```

Db      330  GCCTGCTGACTGGGGGCACTACGGCGGTTTGATGATCGGTATGCTTGGCACTCGCGCTGG 389
QY      339  CACCTTACCTATGATGAGCTGATGGCGTGTGGTGGCGTGTGGCGTCTACAGCGCTTCGGCG 399
Db      390  CACCTTACCTATGCTGATGGCGTGTGGCGGCTGTGTACGAGACGCCGCTTTGCACCT 449
QY      399  TCTCAATAGCTGGCCAGACATGCCATCTGTGATAAGACGACGCTTGTCTTTTGGCCGAT 458
450  GTCATCACTCTGGCCGACACATGATGCCCTGTGATAAGCGCGCTGCTGCTGGCGGAT 509
QY      459  GAAACAAATATGGTGGCAAAATCTCTGGCGGATCTATGATGATCTACAGAAAGCT 518
510  CAAAGAGAGTACGGCAACAAATCAGCTGGCCAGACCTGATGATCTCTGGCTGGCAAGCT 569
QY      519  AGCTCTGGAACCTATGGGCTTTAAACCTTTTGGTTTGCAGGTGGCAGACGAGTATGTC 578
570  GGTCTATGATGATGAGCTTGGCTTACCTCTTACGCGCTCTCTTTGGCGCGCTCGATATTTG 629
QY      579  GCGGCTCGAAGAGCTGATCTGGGAGAGAACCCAGAACCCAGTATGGCTGGGA-----GACAA 632
630  GGAACCCGAAGAATATCTACTGGGGTGCAGAAAGAGTGGCTGGCACTCTCTGACCA 689
QY      633  CGCTATGAGGTGACCCAGAGC-----TCGAAAATCCCTCTGGAGCGCTGACAA 683
690  AGCTTACGGGACGCTGACAAACAGCAGACACATGGAAAACCGCTGGCGGCTGTCCAAT 749
QY      684  GGGACTATCATATGATAACCCGAGACCCAGACCCAGACCCAGACCAAGCACTATCGCTGCTG 743
750  GGGTCTGATCTATGTGACCCCGGAAGGTGTAACGGCCACCTGCTCGCTGAGAACCG 809
QY      744  CGTGTATTCGTGAGACTTTTGGCGGAATGGCAATGATGACAGAGAACCTGTGGTCT 803
810  ACAGAGGTACTTGTAACCTTCCGCCGTATGGCATGAAAGACGAAACCCAGACCT 869
QY      804  CATAGAGGTGACACCTTCCGAAACCACTGCTGCTGCGCATCGCGAGGAATATCT 863
870  CACAGCTGGCGGCCACACCGCTGATTTGTGACGGTATTTGGCAATGCTGCTGCT---T 926
QY      864  GGGCGGACAGCTCTCGCCGCGAGGTATGGAAGAAATGAGCTGGGGTGGAAAACACCTA 923
927  AGCCCTGCCCACAAAGCCTCTGCTGTGAACACAGGGCTTAGTGTGGGCAACCCCA 986
QY      924  CGGACCGGACAGGTGCGGATACCATCAGTGGCTAGAGAGGCGCTGGACCCAGAC 983
987  CATCGAGGCAAGCAAGCAACCGCTGACCTCGGTATCGAAGGTGCTTGGACCCAA 1046
QY      984  CCTTACTCATGGAGCAATAACTTTTGTGAAAACCTCTTTGGTTACGAGTGGAGCTTAC 1043
1047  CCGCAGGAATTCGATTTGGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1106
QY      1044  CAAAGTCCAGGTGAGCTTATCATGTGGAAACCAAGACGCTGCGGGCTGGCAGCAT 1103
1107  AAAGATCTCTGGCGGTGGCCACCTATGGGACCGATTCACATCAAAAAGAAAGAACCA 1166
QY      1104  ACCGGATGCATGATCCAGCAGCTGCGCAGCTCCATTTATGCTCACTACGAGCTGGC 1163
1167  GTTGTGACCCAGCCGCTCTTATTTGCCCAACCGGATCTGACGATCGGATATTTGG 1226
QY      1164  GCTGGCATGACCTGATATAGGAAAATTTCTGCAAGTACTATGAAACCTGATGA 1223
1227  GATTAAGGTAAATCCGAGCTATCGCGCTATCTGCGAATAATTCATGGCCGATCTGATGA 1286
QY      1224  GTTTCAGATGCTTGGGAAAGAGATGTTGCAACTTGACACACAGAGATATGGGACCAA 1283
1287  CTTCAAGAAAGCTTTCCGAGAGCGGTGTTGAGCTGACGACCGCTGACCTGGCGCCGAC 1346
QY      1284  GTTGGGCTTACCTGGGACAGAGTGGCTCGAGAGAGCTCATCTCTGGCAGCCCTATCC 1343
1347  ATCAGGTTTACATCGGCGCGAGAGTGGCGGACAGAGCTCTATTTTGGCAAGACCGATCC 1406
QY      1344  AGATGTAAAGCATCTCTTGTAGAGAAACGATATTGAAGGCTTAAAGCCCAAAATCT 1403
1407  GCGAGGTAC-----ACCGACTCTCTCGAGAGAGTGGCTGACGAGAAATATGCT 1454

```

```

1404 GGAATCGGACTCGCGTAAAGCAGCTGGTAAAGCAGCGATGGCTTCTGCATCTACTTT 1463
1455 ACAAGTGGCTCGGACATAGTAGATGGTCTCCACCGCTTGGACAGTGGCGTACTTAA 1514
1464 TAGAACTCTGACAGAGCGCGGGTGCACAGGTGACAGTATACAGTGGCGCCCAAA 1523
1515 TGGGCTGCTGATGCGGCGGCTGACAGGTGCGCGATTCGCTGGCGCCACAGAA 1574
1524 AGACTGGGACTTAAACACCTCAGCACTTGGCAGGCTGCTCAAACTAGAGAT 1583
1575 CGAGTGGCAGGCAACGAGCGGAGCGCTTGGCGAAAGTCT----- 1616
1584 CCAGGAGAGCTTTAACCAGGCGCATCAGATAACAAAGCAGTATCGTTGGCGAGCTGAT 1643
1617 -----GAGCTCTACGAGAGATCTCTGCCACCGGCGCTAGCATCGGAGGTAT 1670
1644 TGTGCTGGCGCGCTGCGGGTGTAGAAAGCTGCAAAAGATCTGGCATGAGGTGCA 1703
1671 CGTCTGCGCGGTAGCTGAGCATCGAAGACCGGCAAGCAGAGTTCAGATGTGCG 1730
1704 GCGCTCTTCAACCGCGGACGAGCGGATGCCCGCTGAGCAACAGCATGTGGAAGCTTT 1763
1731 CGTCTCTTCCAGCGCGCTGGCATGCGGAGCGGATGCGGAGTACGCGTTCCTT 1790
1764 CGAGCAGCTAGCAGCGGTGAGCGCTTTAGAACTACATTAACCGGAGCTAAGAT 1823
1791 CGCAGCTGCTGAGCGCGCTGGCGGATGGCTTCGCACTGCGCAAGAGAGATATGTGT 1850
1824 ATCCGTGAGAAATGCTCGTAGACGGGCGGAGCTCTGTGCTTTGGCACCAGAAAT 1883
1851 GAAACCGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1910
1884 GACTGCTTTGAGCGGTATGGGTGCTACTGGGACCACTACGACGTTGCGCAGTGG 1943
1911 GACGCTGCTGCTGGCGGTATGCGCTACTGGGACCACTATGCTGGCACCACACGG 1970
1944 AGTGTTCAGAAATAGCGCGGTGAGCTATCCCAATGACTCTTTGTAACCTGCTAGACT 2003
1971 CGTATCCGCTGTGAAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2030
2004 CAAACGCTTAATGGGACCGCGATGATCAGACAAAGTTTTTGAAGCAGAGACTTCAA 2063
2031 GGGGACAGCTGG-----AGCGGTGAGTAGACAGCGCTACGAAATCCGCGACGCA 2084
2064 ACTGGGAGCTAAATGAGGTGACACCGGCTGAGCTGATCTCGGATTCATTCGGA 2123
2085 GACCGGTGCTGAGTGGGCGCTCGCGGTGATCTGGTATTTGGTTCAGCTGCT 2144
2124 GCTAAGAGCTTGCAGAGTGTACGCTGTGCGATCTCGAAGAAAGTTGTTAAGA 2183
2145 ACTCGCTCTTACGACAGAGTGTACCCCGAGGAGGATACCGCGAAGTGTGTCAGAGA 2204
2184 TTTTGTAGAGCTTGGCGCAAGGTATAGGACTTGGAGGCTTTGA 2228
2205 CTGTGCGCGCTTGGACAGAGTATGAAACCGGAGCTTTGCA 2249

```

RESULT 4

```

us-09-940-925A-145
; Sequence 145, Application US/09940925A
; Publication No. 0620030054338A1
; GENERAL INFORMATION
; APPLICANT: BROW, MARY ANN D.
; INVENTOR: LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
; PATHOGENS
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; ADDRESS: 2000 CALIFORNIA STREET, SUITE 2200
; CITY: SAN FRANCISCO

```

```

STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
PUBLICATION DATA: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: PORS-01756
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 145:
SEQUENCE CHARACTERISTICS:
LENGTH: 620 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 145:
us-09-940-925A-145

```

```

Query Match 11.3%; Score 251.8; DB 11; Length 620;
Best Local Similarity 63.9%; Pred. No. 1.6e-71;
Matches 398; Conservative 0; Mismatches 222; Indels 3; Gaps 1;

```

```

QY 916 ACACCTACCGCAGCGGACAGCGGTACCATCACCAGTGGAGCGCTGATGG 975
DB 1 AGCTGTATGGCAGCGGACAGCGGTAAAGGAGCGGATCACCAGCGCATCGATGG 60
QY 976 ACAGAGCCCTACTACTGAGGAGCAATTTTTTGAACACCTTTTGGTTAGGAGTG 1035
DB 61 AGAACACCCCGACGAATGGGACACAGTTTCTCGAGTCTGAGGCTAGAGAGTG 120
QY 1036 GAGCTTACCAAAATGTCAGCTGGAGCTTATCAGTGGNACCAAGACGTCGGCGCT 1095
DB 131 GAGCTGACCAAGAGACCTCTGCTGGCGTTTGCAATACACCGCAAGAGCGCGGTGCC 180
QY 1096 GGCACATACCGGATGACATGATCCACAGCAAGTGCAGCTCCATTATTCCTACAGT 1155
DB 181 GACACATCGCGGCGCTTGG-----CGGCGCGCGGCGCTCCCGAGGATGCTGGCCT 237
QY 1156 GACCTGGCGCTGGGAGTGGAGCGCTGATACGAAATTTCTCGAGCTACTATGAAGC 1215
DB 238 GACCTCTCGCTGGGCGTGGATCCGATCTATGAGGAGCATCGCTGCTGGTGAACAC 297
QY 1216 CTGATGAGTTTTCAGATGCTTTCGGGAAGCATGGTACAACTGACACACAGAGATATG 1275
DB 298 CCGGAGAAATGTCGAGTGGCGAGGCTTGGTCAAGTCAAGTCAAGTCAAGTCAAG 357
QY 1276 GAGCAAGATGGCGCTACTCTGGACAGCAAGTGCCTCAGAGAGCTCATCTGCGAAGC 1335
DB 358 GGTCCCTTGGGAGATACCTTGGGCGGTGGTCCCGACAGAGACCTTGTGTGAGAT 417
QY 1336 CCTATACAGATGTAGCATCTCTTCTAGACGAAAGAGATATTGAAGGCTTAAAGCC 1395
DB 418 CGGTCCTTGGGTGACGACCGCTCTGGGAGCGGAGATTCGACGCTTAAAGC 477
QY 1396 AAAATCTGGANTCGGAGCTGAGGTGAGGAGTGTGTGAAGCAGCGATGGCTTCTGCA 1455
DB 478 CAGATCTCGCATGGGATGACTGTCTCACAGCTATTTTCGACCGCATGGCGGCGCG 537
QY 1456 TCTACTTTTGAAGACTCTGACAGCGCGGCGGTGCCAAGCTGCGACTATACAGTGGCC 1515
DB 538 TCGTGTGCTGCTGGTGGACAGACCGCGGCGCGCAAGCTGCTGGTGGTGGTGGTGG 597

```

QY 1516 CCACAAAAGCTGGGAAGTAAA 1538  
 Db 598 CCACAAAGCTGGGCGGAGGTCAA 620

## RESULT 5

US-09-940-925A-149/c  
 ; Sequence 149, Application US/09940925A  
 ; Publication No. US2003005438A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LYMICHEV, VICTOR I.  
 ; INVENTOR: LYMICHEV, VICTOR I.  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CARROLL, PETER G.  
 ; REGISTRATION NUMBER: 32,837  
 ; REFERENCE/DOCKET NUMBER: FORS-01756  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 149:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 620 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 149:

## TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

PATHOGENS

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESS: CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/940,925A

FILING DATE: 10-Jun-2002

CLASSIFICATION: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: CARROLL, PETER G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: FORS-01756

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 149:

SEQUENCE CHARACTERISTICS:

LENGTH: 620 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 149:

US-09-940-925A-149

Query Match

Best Local Similarity 11.3%; Score 251.8; DB 11; Length 620;

Matches 398; Conservative 0; Mismatches 222; Indels 3; Gaps 1;

QY 916 AACACATACGGCCCGGACAGCTCCGGATACCATCAGATGAGCTAGAGCGCGCTGG 975

Db 620 AGCTCGTATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 561

QY 976 ACCAGACCCCTACTCAATGAGACACACTCTTTTGAACACTCTTTGTTAGAGTGG 1035

Db 560 AGACACACCCCGGAAATGGGACACACTCTTCTCGAGATCCTGTAGCGCTACGAGTGG 501

QY 1036 GAGCTTACCAAAAGCTCAGCTAGCTATCATCTCGAAGAACAAAGAGCGCGCGGCT 1095

Db 500 GAGCTGACGAGAACCCCTGCTGGCTTTGGCATACACCGGACGAGCGCGCGCGGCT 441

QY 1096 GCACCATATCATCCGATGACATGATCCACGACGCGACGCTCCATTTATGCTACATCG 1155

Db 440 GCACCATATCCCGGACCCCTTCG---GCGGCGCCAGGCGGCTCCCGGACGATGCTGGCCACT 384

QY 1156 GAGCTCGGCGCTCCGCAATGACCTGATTAACGAAAAATTTCTCGAGCGTACTATGAAAC 1215

Db 383 GAGCTCTGCTCGCGGTGGATCCGATCTATGCGGATACAGGCTGCTGGCTGGACAC 324

QY 1216 CCTGATGAGTTTGACAGATGCTTTTCGCGAAGCATGTGTAACATGACACAGAGATG 1275  
 Db 323 CCACGAGAAATGGCGGACGAGTTCCGCAAGCGCTGTACAGCTGATCCACGACAGTG 264  
 QY 1276 GGACAAAGAGTCCCTAGCTCGGACGACGAGTGGCTGTAGGAGAGCTCATCTCGGAGAC 1335  
 Db 263 GGTCCCGTTGCGAGATACCTTTTGGCCGCTGTCCCAAGCAGACCTCTGTGTGCGCAGAT 204  
 QY 1336 CCTATACCAAGATGATAGCATCTCTGTAGACGAAAAAGATATTGAAGCGCTTAAAGCC 1395  
 Db 203 CGGTCCTCTCGGCTGACGACGACCTGTCTGGGAGAACCGAGATGCGACGCTTAAAGC 144  
 QY 1396 AAAATCCCTCGAATCGGAGCTGACGTAAGCAGCTGTATAGCAGCGCTGCTTCTCA 1455  
 Db 143 CAGATCTGCGATCGGATGACTGTCTCAGACTAGTTTTCGACGCGCATGGCGCGCGG 84  
 QY 1456 TCTACTTTTGAAGAACTCTGACAAAGCGCGCGGTGCCAACGGTGCAAGTATACGATGGCC 1515  
 Db 85 TCGCTGCTTCCTGCTGTGTCAGCAAGCGCGCGCGCGCCCAACGCTGTGCGATCCGCTCGAG 24  
 QY 1516 CCACAAAGCTGGGAGTAA 1538  
 Db 23 CCACAAAGCTGGGAGGTCAA 1

## RESULT 6

US-09-941-193A-145  
 ; Sequence 145, Application US/09941193A  
 ; Publication No. US20030108873A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BROW, MARY ANN D.  
 ; INVENTOR: LYMICHEV, VICTOR I.  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CARROLL, PETER G.  
 ; REGISTRATION NUMBER: 32,837  
 ; REFERENCE/DOCKET NUMBER: FORS-01756  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 145:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 620 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 145:  
 ; US-09-941-193A-145

## TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

PATHOGENS

NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESS: CARROLL

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/941,193A

FILING DATE: 28-Aug-2001

CLASSIFICATION: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: CARROLL, PETER G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: FORS-01756

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 145:

SEQUENCE CHARACTERISTICS:

LENGTH: 620 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 145:  
 ; US-09-941-193A-145

Query Match

Best Local Similarity 11.3%; Score 251.8; DB 11; Length 620;

Matches 398; Conservative 0; Mismatches 222; Indels 3; Gaps 1;



```

Qy 916 AACACCTACGCCACCGCACCGTGGGATACCATCACTGACATAGAACGCGCTG 975
Db 1 ACCTCGTATGCGACCGGACCGGACCGTANGAGCGGATCCTCGGCATCGAGTGTATG 60
Qy 976 ACCAAGACCCCTACTCAATGAGCAATATCTTTTAAACACTCTTTGGTTACGAGTGG 1035
Db 61 ACAGAACCCCGAGCAATGAGGACACAGTTTCTCGAGATCTGTAACGCTACGAGTGG 120
Qy 1036 GAGCTTACAAAGTCCGACTGAGAGCTTATCATGTTGGAACCAAAAGCGGCGGGGT 1095
Db 121 GAGCTGAGGAGAGCCCTCTGGCGTTGGGATACACCCAGCAAGAGCGGCGCGGTGC 180
Qy 1096 GGCACCATATCGAGTGCACATGATCAGCAGATGCGACGCTCCATTTATCTCACTAG 1155
Db 181 GGCACCATATCGAGTGCACATGATCAGCAGATGCGACGCTCCATTTATCTCACTAG 1155
Qy 1156 GAGCTGGGCTGGGAGCGCTGATGAGCAAAATTTCTGAGGACTATGAAAC 237
Db 238 GAGCTTCTGCTGGGTGGATCGATCATGACGGGATCAGCGGTGCTGGCTGCAAC 295
Qy 1216 CTTGATGAGTTTGCAGATCTCTTTGTAAGCAAAAGATATTAAGGCTTAAAGGCC 1275
Db 298 CCGGAGGAATTTGGCGAGCTGTCGCGAGCTGGTATGAGCTGATTCACCGAGAGATG 357
Qy 1276 GAGCAAAAGCGGCTGACTGCGACAGAGAGTGGCTTCAAGAGAGCTCATCTGGCAAG 1335
Db 358 GGTCCGCTTGGAGATACCTTTGGGCGCTGGTCCCAAGCAGAGCCCTGCTGGCAGAT 417
Qy 1336 CTTATACAGATGTAGCCATCTCTTTGTAAGCAAAAGATATTAAGGCTTAAAGGCC 1395
Db 418 CCGGCTGCTGGCTGATCAGCAGCTCTGTCGCGAAGCGGAGATTGCGAGCTTAAAGAG 477
Qy 1396 AAATCTCTGAGTCCGAGTGGAGTGAAGCGAGTGGTCAAGCGGAGTGGCTTGAAG 1455
Db 478 CAGATCTGCGATCGGATGACTGTCTCAGCAGTAGTTTCAGCGCATGGGCGCGGCG 537
Qy 1456 TCTACTTTTAAAGACTCTGACAGCGCGCGGTGCGCAACCTATACAGCTGGCC 1515
Db 538 TGGTGTCTGCTGGTACGACAGCGCGCGCGCCACAGTGGTGTGCATCCGCTCGAG 597
Qy 1516 CCNCAAAAGACTGGGAATAA 1538
Db 598 CCNCAAGTGGGTGGAGGTCAA 620

```

## RESULTS 7

US-09-941-193A-149/C

; Sequence 149, Application US/09941193A

; Publication No. US20030108873A1

; GENERAL INFORMATION:

; APPLICANT: BROW, MARY ANN D.

; LYANICHEV, VICTOR I.

; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

; VIRUSES

; NUMBER OF SEQUENCES: 163

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN &amp; CARROLL

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION NUMBER:

; APPLICATION NUMBER: US/09/941,193A

; PUBLICATION DATE: 08-Aug-2001

; CLASSIFICATION:

; CLASSIFICATION: G06F000378&gt;

; ATTORNEY/AGENT INFORMATION:

;

;

;

NAME: CARROLL, PETER G.

REGISTRATION NUMBER: 32,837

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 149:

SEQUENCE CHARACTERISTICS:

LENGTH: 620 base pairs

TYPE: nucleic acid

STR: 5' to 3' direction

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 149:

US-09-941-193A-149

Query Match 11.3% Score 251.8; DB 11; Length 620;

Sequence Similarity 63.94; Pct 1.66; Mismatches 222; Indels 3; Gaps 1;

Matches 395; Conservative 0; Mismatches 222; Indels 3; Gaps 1;

```

Qy 916 AACACCTACGCCACCGCACCGTGGGATACCATCACTGACATAGAACGCGCTG 975
Db 620 ACCTCGTATGCGACCGGACCGGACCGTANGAGCGGATCCTCGGCATCGAGTGTATG 561
Qy 976 ACCAAGACCCCTACTCAATGAGCAATATCTTTTAAACACTCTTTGGTTACGAGTGG 1035
Db 620 ACCTCGTATGCGACCGGACCGGACCGTANGAGCGGATCCTCGGCATCGAGTGTATG 561
Qy 976 ACAGAACCCCGAGCAATGAGGACACAGTTTCTCGAGATCTGTAACGCTACGAGTGG 501
Db 560 GAGCTTCTGCTGGGTGGATCGATCATGACGGGATCAGCGGTGCTGGCTGCAAC 295
Qy 1036 GAGCTTACAAAGTCCGACTGAGAGCTTATCATGTTGGAACCAAAAGCGGCGGGGT 1095
Db 500 GAGCTTCTGCTGGGTGGATCGATCATGACGGGATCAGCGGTGCTGGCTGCAAC 295
Qy 1096 GGCACCATATCGAGTGCACATGATCAGCAGATGCGACGCTCCATTTATCTCACTAG 1155
Db 440 GGCACCATATCGAGTGCACATGATCAGCAGATGCGACGCTCCATTTATCTCACTAG 1155
Qy 1156 GAGCTGGGCTGGGAGCGCTGATGAGCAAAATTTCTGAGGACTATGAAAC 1215
Db 383 GAGCTTCTGCTGGGTGGATCGATCATGACGGGATCAGCGGTGCTGGCTGCAAC 324
Qy 1216 CTTGATGAGTTTGCAGATCTCTTTGTAAGCAAAAGATATTAAGGCTTAAAGGCC 1275
Db 323 CCGCAGGAATTTGGCGAGAGTTCGCCAAGGCTTGTACAAAGTGTATCCACGAGCATC 264
Qy 1276 GGACAAAGTGGCTACTTGGGACGAGAGTGGCTCAGGAGACTCATCTGGCAGAC 1335
Db 263 GGTCCCTCGGAGATACCTTGGCGGCGGTGTCGCCAAGCAAGCATTTGCTGTGCGAGAT 204
Qy 1336 CTTATACAGATGAAGCATCTCTTCTAGACCAAAAGATATTAAGGACTTAAAGGCC 1395
Db 203 CGGTCTCTCTGGTTCAGCAGACCTCTCTGGGAGCGGAGATTTGCAGGCTTTAAGAC 144
Qy 1396 AAATCTCTGAGTCCGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGG 1455
Db 143 CAGATCTGGGATCCGGATTGACTTCTACAGACTATTTGACACCGCTGGGCGGCGG 84
Qy 1456 TCTACTTTTAAAGACTCTGACAGCGCGGCGGTGCGCAACGCTATTAAGGACTTAAAGGCC 1515
Db 83 TGGTGTCTCGTGGTGAAGCAAGCGCGGCGGCGCAACGCTGCTGCATCGCTCGAG 24
Qy 1516 CCNCAAAAGACTGGGAATAA 1538
Db 23 CCNCAAGTGGGTGGAGGTCAA 1

```

## RESULT 8

US-09-940-925A-143

; Sequence 143, Application US/09940925A

; Publication No. US20030054338A1

; GENERAL INFORMATION:

; APPLICANT: BROW, MARY ANN D.

; LYANICHEV, VICTOR I.

; OLIVE, DAVID M.

;

;

;

;

;

[illegible][illegible]

```

Db 181 GGCACCTCCGGACCGCTTGC---GCGGCGCAGCGGCTCCGCGAGCATCTGGCCCT 237
Qy 1156 GACCTGGGCTGGCGATGACCTGATTTAGGAAAAATTTCTCGAGGCTACTATGAAC 1215
Db 238 GACCTCTCGCTGGGCTGGATGCGGATCTATGAGCGGATCACCGCTCGCTGGCTGGAAC 297
Qy 1216 CCGCATGATTTGCGAGTCTTTTCGGGAAGGATGCTACAACTGTAACAGACAGATG 1275
Db 298 GCGGCTGGGCTGGGATGCTGCGAAGGCTGGTGAAGCTGATTCGACGAGGATGAT 357
Qy 1276 GGAACAAAGGTGGCTACTCTGGGACAGAGGCTCTGAGGAAGACTCATCTGGAAGAC 1335
Db 358 GGTCCGCTGGGATGATCTCTGGGCGGCTGTCGCAAGCAGCAGCCCTCTCTGGCAGAT 417
Qy 1336 CCGTATACAGATGATGAGCATCTCTGTGTAAGCAAAAGATATGAGGCTTAAAGCG 1395
Db 418 CCGGCTCGCTGGCTGACGACGAGCTCTGGCGAGCGGATGCGACCTTTAAGGC 477
Qy 1396 AAAAACTCGAATCGGATCGAGGTAAAGCTGTAAGCAGCGCTGTAAGCAGGCTTCTGCA 1455
Db 478 CAGATCTGCGCATCGGATGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 537
Qy 1456 TCGATCTTGAAGCTATGCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
Db 538 TCGCTGCTGGGTAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 620
Qy 1516 CCACAAAGACTGGGAAGTAA 1538
Db 598 CCACAAAGCTGGGTGGAGTCAA 620

```

## RESULT 10

```

US-09-940-925A-147/c
; Sequence 147, Application US/09940925A
; Publication No. US20030054338A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; CONTACT: LYMICHEV, VICTOR I.

```

## TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

```

; PATHOGENS

```

```

; NUMBER OF SEQUENCES: 165

```

```

; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: MEDLEN & CARROLL

```

```

; STREET: 220 MONTGOMERY STREET, SUITE 2200

```

```

; CITY: SAN FRANCISCO

```

```

; STATE: CALIFORNIA

```

```

; COUNTRY: UNITED STATES OF AMERICA

```

```

; ZIP: 94104

```

```

; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; APPLICATION NUMBER: US/09/940,925A

```

```

; FILING DATE: 10-Jun-2002

```

```

; CLASSIFICATION: <Unknown>

```

```

; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: CARROLL, PETER G.

```

```

; REGISTRATION NUMBER: 32,497

```

```

; REFERENCE/DOCKET NUMBER: FORS-01756

```

```

; TELEPHONE: (415) 705-8410

```

```

; TELEFAX: (415) 397-8338

```

```

; INFORMATION FOR SEQ ID NO: 147:

```

```

; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 620 base pairs

```

```

; TYPE: nucleic acid

```

```

; STRANDEDNESS: double

```

```

; ANALYSIS: BLAST

```

```

; MOLECULE TYPE: DNA (genomic)

```

```

; US-09-940-925A-147:

```

```

Query Watch 11.2s; Score 250.2; DB 11; Length 620;
Best Local Similarity 63.7%; Pred. No. 5.4e-71;
Matches 397; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

```

```

Qy 916 AACACCTACGGCACCGGACGCGGATACCATCAGCTAGTGCAGCTAGAAGCGCCCTGG 975
Db 916 AACACCTACGGCACCGGACGCGGATACCATCAGCTAGTGCAGCTAGAAGCGCCCTGG 975
Qy 976 ACCAAGCTCTGACTCAATGAGCACAATCTTTTGAAGAACTCTTTGGTACAGATGG 1035
Db 976 ACCAAGCTCTGACTCAATGAGCACAATCTTTTGAAGAACTCTTTGGTACAGATGG 1035
Qy 560 AGAACAACCCGACCAATAGGGACACAGTTTTCCTCGAGATCTCTACCGCTACAGATGG 501
Db 560 AGAACAACCCGACCAATAGGGACACAGTTTTCCTCGAGATCTCTACCGCTACAGATGG 501
Qy 1036 GAGCTTACCAAAAGTCCAGCTGAGGATTCATGATGGAAGCAAAAGACGCTGCCGGGCT 1095
Db 1036 GAGCTTACCAAAAGTCCAGCTGAGGATTCATGATGGAAGCAAAAGACGCTGCCGGGCT 1095
Qy 500 GAGCTGAGGAGAGAGGCTCTGTCGCTTTGGCAATACACCGCCAAAGACGCGCCGATGCC 441
Db 500 GAGCTGAGGAGAGAGGCTCTGTCGCTTTGGCAATACACCGCCAAAGACGCGCCGATGCC 441
Qy 1096 GGCACCATACCGATGACATGATCAGCAAGTCCGACGCTCATATTATGCTCTCACTAG 1155
Db 1096 GGCACCATACCGATGACATGATCAGCAAGTCCGACGCTCATATTATGCTCTCACTAG 1155
Qy 440 GGCACCATACCGACCGCGCTGCG---GCGGCGCAGGCGCTCCCGACGATCTGCGCACT 384
Db 440 GGCACCATACCGACCGCGCTGCG---GCGGCGCAGGCGCTCCCGACGATCTGCGCACT 384
Qy 1156 GACCTGGGCTGGCGATGACCCCTGATTAGAAAAATTTCTCGACGCTACTATGAAC 1215
Db 1156 GACCTGGGCTGGCGATGACCCCTGATTAGAAAAATTTCTCGACGCTACTATGAAC 1215
Qy 384 GACCTCTGCTGGGATGATCTGATCTGAGCGGTACGCGTGCCTGCTGGACAC 324
Db 384 GACCTCTGCTGGGATGATCTGATCTGAGCGGTACGCGTGCCTGCTGGACAC 324
Qy 1216 CCGTATACAGATGTAAGCCATCTCTGTAGACGAAACAGATATTGAAGCGCTTAAAGCC 1395
Db 1216 CCGTATACAGATGTAAGCCATCTCTGTAGACGAAACAGATATTGAAGCGCTTAAAGCC 1395
Qy 203 CCGGCTCCCTCGGCTCAGCCACGACCTCTCGCGAAGCGGAGATTGCCAGGCTTAAGAC 144
Db 203 CCGGCTCCCTCGGCTCAGCCACGACCTCTCGCGAAGCGGAGATTGCCAGGCTTAAGAC 144
Qy 1396 AANATCTGATGGGATGAGCGGTAAAGGAGCTGTAGCAGCAGCTATCAGCGATGGGCTTGA 1455
Db 1396 AANATCTGATGGGATGAGCGGTAAAGGAGCTGTAGCAGCAGCTATCAGCGATGGGCTTGA 1455
Qy 143 CAGATCGGGCATCCGGATGACTGTCTCAGACGCTATTTCCAGCCGATTCGGCGGCG 84
Db 143 CAGATCGGGCATCCGGATGACTGTCTCAGACGCTATTTCCAGCCGATTCGGCGGCG 84
Qy 1456 TCTACTTTTGAACCTCTGACAAAGCGCGCGCTGCCACGCTGCACGTATACGCTGGCC 1515
Db 1456 TCTACTTTTGAACCTCTGACAAAGCGCGCGCTGCCACGCTGCACGTATACGCTGGCC 1515
Qy 83 TCGCTGCTCGGTGTAGCAGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 84
Db 83 TCGCTGCTCGGTGTAGCAGACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 84
Qy 1516 CCACAAAGACTGGGAAGTAA 1538
Db 23 CCACAAAGCTGGGTGGAGTCAA 1

```

## RESULT 11

```

US-09-940-925A-150/c
; Sequence 150, Application US/09940925A
; Publication No. US20030054338A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.

```

```

; CONTACT: LYMICHEV, VICTOR I.

```

```

; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

```

```

; PATHOGENS

```

```

; NUMBER OF SEQUENCES: 165

```

```

; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: MEDLEN & CARROLL

```

```

; STREET: 220 MONTGOMERY STREET, SUITE 2200

```

```

; CITY: SAN FRANCISCO

```

```

; STATE: CALIFORNIA

```

```

; COUNTRY: UNITED STATES OF AMERICA

```

```

; ZIP: 94104

```

```

; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/940,925A
: FILING DATE: 10-Jun-2002
: CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: CARROLL, PETER G.
: REGISTRATION NUMBER: FORS-01756
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 150:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 820 base pairs
: TYPE: DNA (genomic)
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 150:
US-09-940-925A-150

```

```

Query Match      11.2%; Score 250.2; DB 11; Length 620;
Best Local Similarity 63.7%; Pred No. 5.4e-71;
Matches 397; Conservative 63; Mismatches 223; Indels 3; Gaps 1;

QY 916 AACACCTACGCGACCGGACGCTGGCGGAFACCATCAGCTAGAGCGGCGCTGG 975
Db 620 AGCTGCTATGGCCGAGCCGACGAGCGCATACCCGAGGCTGATGTTG 561
QY 976 ACCAGACCCCTACTCATGTGACGACATACTTTTGAAGACCTCTTGTACGAGTGG 1035
Db 560 ACGAACACCCGAGAAATGGGACACAGTTTCTCTCGAGATCTGTACGGCTACGAGTGG 501
QY 1036 GAGCTTACCAAAATCCGAGCTGAGCTTATCAGTGGAAACCAAAAGACGGTCCGGGGCT 1095
Db 500 GAGCTTGACGAGAGCCCTGCTGGCTGTGGCATACACCGCCAGAGCGCGCGGTGCC 441
QY 1095 GGCACCATCCGAGATGCATATGCCAGCAAGTCGACGCTCCATTTATGCTCACTAGG 1155
Db 440 GGCACCATCCCGGACCCGCTTGC---GCGGGCAGGGGCTCCCGGACGATGCTGCCACT 384
QY 1156 GAGCTGGCGCTGCCATGACCCCTGATTACGAAAAATTTCTCGAGGTACTATGAAC 1215
Db 383 GAGCTTCCGCTCCGGCTGGATTCGATCTATGCGCGATACGCTGCTGCTGGGACAC 324
QY 1215 CCTGATGATGTTGACGATGCTTTTGGGAAAGCATGTGTACAACTGACACACAGAGATG 1275
Db 323 CCGGAGGAGATTGGCCGACGAGTTCGCGAAGCCCTGGTCAAGCTGATCAACCGAGACATG 264
QY 1276 GGACCAAAAGGCTGCTACTTGGGACACAGAGTGCCTCAGSAGAACCTCATCTGGCAAGC 1335
Db 263 GGTCCGCTGTGATGATCTTGGGCGCTGTGCTCCCAAGACCCCTGCTGGGAGAT 204
QY 1336 CCTATACACAGATGTATGACATCTCTTTGTAGACGAAAGCATGATTGAGCGCTAAAGCC 1395
Db 203 CCGGCTCCCTGGGTGACGACAGACAGCTCTGTCGGGAAAGCCGAGATGCCAGCTTTAAGAC 144
QY 1396 AAATCTCTGAAATGGGACTGACAGGTAAGGACGCTGTGTAGAGAGCCATGGGCTTCTGCA 1455
Db 143 CAGATCTCTGCGATCGGATTTGATCTCTACACTACTTGTTCAGCCGATGGGCGGGG 84
QY 1456 TCTACTTTTGAAGAACTCTTGACAAAGCGCGCGGTGCAACCGGTGCAGTATAGCACTGCC 1515
Db 83 TGTCTGCTTCGCTGGTATAGGACAAAGCGCGCGCGGCAACCGGTGATGCTATCGGCTGCAG 24
QY 1516 GCACAAAAGATCGGAGTMAA 1538
Db 23 CCAAGATTCGGGTGGGTCCTAA 1

```

RESULT 12

```

US-09-941-193A-143
: Sequence 143: Application US/09941193A
: FILING DATE: 28-Aug-2001
: GENERAL INFORMATION:
: APPLICANT: BROW, MARY ANN D.
: LYANICHEV, VICTOR I.
: OLIVE, DAVID M.
: TITLE OF INVENTION: PATH DETECTION AND IDENTIFICATION OF
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
: COMPUTER PROGRAM: PC-DOS/MS-DOS
: MEDIUM TYPE: 5.25 floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/941,193A
: FILING DATE: 28-Aug-2001
: CLASSIFICATION: <Unknown>
: ATTORNEY/AGENT INFORMATION:
: NAME: CARROLL, PETER G.
: REGISTRATION NUMBER: 32,837
: REFERENCE/DOCKET NUMBER: FORS-01756
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 143:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 620 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: SEQUENCE DESCRIPTION: SEQ ID NO: 143:
US-09-941-193A-143

```

```

Query Match      11.2%; Score 250.2; DB 11; Length 620;
Best Local Similarity 63.7%; Pred No. 5.4e-71;
Matches 397; Conservative 63; Mismatches 223; Indels 3; Gaps 1;

QY 916 AACACCTACGCGACCGGACGCTGGCGGAFACCATCAGCTAGAGCGGCGCTGG 975
Db 1 AGCTGCTATGGCCGAGCCGATGAGACGGATCCAGCGCATCAGCGCTAGTCTA 60
QY 976 ACCAAGACCCCTACTCATGTGACGACATAACTTTTTTGAAGACCTCTTTGTTACGAGTGG 1035
Db 61 ACGAACACCCGAGAAATGGGACAGATTTCTCTCGAGATCTCTGTACGGCTACGAGTGG 120
QY 1036 GAGCTTACCAAAATCCGAGCTGAGCTGATTCAGTGGAAACCAAAAGACGGTCCGGGGCT 1095
Db 121 GAGCTGACGAGAGCCCTGCTGGCGCTTGGCAATACACCGCACGACGCGCGCGTGGC 180
QY 1096 GGACCAATACGAGATGCATGATCCGAAAGTCGACGGCTCCATTTATGCTCACTAGG 1155
Db 181 GGACCAATCCCGAGCCGCTTGC---GCGGGCAGGGGCTCCCGGACGATGCTGCCACT 237
QY 1156 GAGCTGGCGCTGGCATGAGCCCTGATGACGAAATTTCTGAGCGCTACTATGAAC 1215
Db 238 GAGCTTCTGCTGGGCTGATCCGATCTATAGCGGATCAGCGTGCCTGGCTGGAAC 297
QY 1216 CCTGATGATGTTGAGATGCTTTCGGGAAAGCATGTTACAAATCTGACACACAGAGATG 1275
Db 298 CCGGAGGAAATTTGGCCGACGAGCTGTGCAAGGCTTGTCAAGCTGTCAACCGAGACATG 357
QY 1276 GGACCAAAAGGCTGCTACTTGGGACCAAGAGTGCCTCAGAGAGCTCATCTTGGCAGAC 1335

```

Db 358 GGTCCGCTTGGAGATACCTTGGCGCGTGTGCCCAAGCAGACGCTGCTGGCAGAT 417  
 Qy 1336 CCFATACAGATGATGACCTCTCTGTFAGCCGCAAAACGATATGAGCGCTTAAGGC 1395  
 Db 418 CGGTGCTCTGGCTCGCCAGCAGCTGTGCGCGAAGCGGAGATTGCCAGCTTAAGGC 477  
 Qy 1396 AAATCTCGAATCGGATCGGATCGAGTAAAGGATGGTAAAGCAGCGCATGGCTTCTGCA 1455  
 Db 478 CAGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGG 537  
 Qy 1456 TCTACTTTTGAACCTCTGCAAGCGCGCGTGGCCACGCGTGCAGCTGATACGATCGGC 1515  
 Db 538 TGGTCTGCTCGTGTAGCAGACAGCGCGCGCGCCCAAGCGTGTGCTGCTGCTGCTG 597  
 Qy 1516 CCACAAAGAGCTGGGAAGTAA 1538  
 Db 598 CCACAAAGCTGGGAGGTCAA 620

RESULT 13  
 US-09-941-193A-146  
 ; Sequence 146, Application US/09941193A  
 ; Publication No. US20030108873A1  
 ; GENERAL INFORMATION: MARY ANN D.  
 ; APPLICANT: LYAMICHEV, VICTOR I.  
 ;  
 ; OLIVE, DAVID M.  
 ; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF  
 ;  
 ; NUMBER OF SEQUENCES: 165  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MEDLEN & CARROLL  
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200  
 ; CITY: SAN FRANCISCO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: UNITED STATES OF AMERICA  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/941.193A  
 ; FILING DATE: 28-AUG-2001  
 ; CLASSIFICATION: <GKHOWD>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CARROLL, PETER  
 ; REGISTRATION NUMBER: 32,837  
 ; REFERENCE/DOCKET NUMBER: FORS-01756  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 397-8338  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 146:  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 146:  
 ; LENGTH: 620 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 146:  
 US-09-941-193A-146

Query Match 11.28; Score 250.2; Db 11; Length 620;  
 Best Local Similarity 63.78; Pred. No. 5.4e-71;  
 Matches 397; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

Qy 916 AACCTCTAGCCCGGACCGGATGATGATGATGATGATGATGATGATGATGATGATG 975  
 Db 1 AGCTCGTATGTCACCGGACCGGATGATGATGATGATGATGATGATGATGATGATG 60  
 Qy 976 ACCAAGCGGCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1035

Db 61 ACGAACACCCCGAATAATGGGACAAACAGTATTCTTCGAGATCTCTGACGGCTACGAGTGG 120  
 Qy 1036 GAGCTTCAAAAGTTCACCTGCGAGCTGATATCATGTTGGAACCAAAAGACGGTGGCGGGCT 1095  
 Db 121 GAGCTGACGAGAGCCCTCTGGCGCTTGGCAGTACACCGACGACGACGCGCGCGTCC 180  
 Qy 1096 GGCACATACGGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1155  
 Db 181 GGCACATACGGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 237  
 Qy 1156 GAGCTGCGGCTGCGATGAGCAGCTGATATGCAAAAATTTTCGAGCGGTACTATGAAAC 1215  
 Db 238 GAGCTTCTGCTCGGGTGGATCGATCTATGAGCGGATCATGCGTGGCTGGTGAACAC 297  
 Qy 1216 CTTGATGAGTTTCGAGATGCTTTTGGCAAGAGATGGTACAAACTGACACACAGATATG 1275  
 Db 298 CCGGAGAAATTTGGCCGACGAGTTGCGCAGGCTGGTCAAGCTGATCCGCGACATG 357  
 Qy 1276 GGACCAAGGTGCTACTCTGGGACGAGATGGCTTCAGGAGACCTCATCTGCGCAAGC 1335  
 Db 358 GGTCCGCTTGGAGATACCTTGGCGCGTGTCCCAAGCAGACGCTGCTGTGCGAGAT 417  
 Qy 1336 CCTATACAGATTAAGCATCTCTTGTAGACGAAACGATATTGAAGGCTCAAAAGCC 1395  
 Db 418 CGGCTCTTGGCTGCGCAGCAGCTTGTGCGGACGCGGAGATTTCGAGCTTAAGGC 477  
 Qy 1396 AAATCTCGAATCGGATCGGATCGAGTAAAGGATGGTAAAGCAGCGCATGGCTTCTGCA 1455  
 Db 478 CAGATCTGCGATCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 537  
 Qy 1456 TCTACTTTTGAACCTCTGCAAGCGCGCGTGGCCACGCGTGCAGCTGATACGATCGGC 1515  
 Db 538 TGGTCTGCTCGTGTAGCAGACAGCGCGCGCCCAAGCGTGTGCTGCTGCTGCTG 597  
 Qy 1516 CCACAAAGAGCTGGGAAGTAA 1538  
 Db 598 CCACAAAGCTGGGAGGTCAA 620

RESULT 14  
 US-09-941-193A-147/G  
 ; Sequence 147, Application US/09941193A  
 ; Publication No. US20030108873A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BROW, MARY ANN D.  
 ; OLIVE, DAVID M.  
 ; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF  
 ;  
 ; NUMBER OF SEQUENCES: 165  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MEDLEN & CARROLL  
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200  
 ; CITY: SAN FRANCISCO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: UNITED STATES OF AMERICA  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/941.193A  
 ; FILING DATE: 28-AUG-2001  
 ; CLASSIFICATION: <GKHOWD>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CARROLL, PETER  
 ; REGISTRATION NUMBER: 32,837  
 ; REFERENCE/DOCKET NUMBER: FORS-01756  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 147:  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 147:  
 ; LENGTH: 620 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 147:  
 US-09-941-193A-147

Query Match 11.28; Score 250.2; Db 11; Length 620;  
 Best Local Similarity 63.78; Pred. No. 5.4e-71;  
 Matches 397; Conservative 0; Mismatches 223; Indels 3; Gaps 1;

Qy 916 AACCTCTAGCCCGGACCGGATGATGATGATGATGATGATGATGATGATGATGATG 975  
 Db 1 AGCTCGTATGTCACCGGACCGGATGATGATGATGATGATGATGATGATGATGATG 60  
 Qy 976 ACCAAGCGGCTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1035



Of 1516 CCACAAAGACTCGGAGCTAAA 1538  
Db ||||| | ||||| ||  
23 CCACAGTCGGTGGGAGGTCAA 1

Search completed: October 8, 2003, 19:55:28  
Job time : 577.431 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2003, 07:49:13 ; Search time 4638.63 Seconds  
(without alignments)  
11726:187 Million cell updates/sec

Title: US-09-884-889-7

Perfect score: 2238

Sequence: 1 atgaataacacacactc.....acgggttgatctgaataa 2238

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_esthum:\*
- 4: em\_esthum:\*
- 5: em\_estcov:\*
- 6: em\_estpro:\*
- 7: em\_estpro:\*
- 8: em\_estpro:\*
- 9: gb\_estcl:\*
- 10: gb\_estcl:\*
- 11: gb\_estcl:\*
- 12: gb\_estcl:\*
- 13: gb\_estcl:\*
- 14: gb\_estcl:\*
- 15: em\_estfun:\*
- 16: em\_estfun:\*
- 17: em\_estfun:\*
- 18: em\_estfun:\*
- 19: em\_estfun:\*
- 20: em\_estfun:\*
- 21: em\_estfun:\*
- 22: em\_estfun:\*
- 23: em\_estfun:\*
- 24: em\_estfun:\*
- 25: em\_estfun:\*
- 26: em\_estfun:\*
- 27: em\_estfun:\*
- 28: gb\_estcl:\*
- 29: gb\_estcl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Length	ID	Description
1	296.2	13.2	786	28 BH403595
2	269.1	12.6	182	18 AF075865
3	269.1	12.6	182	18 AF075865
4	261	11.7	565	28 BH401331

#### ALIGNMENTS

RESULT	Accession	Length	Score	Definition	Accession	Length	Score	Definition
1	BH403595	786	296.2	AG-ND-127D17, AF-ND-TAM Anopheles gambiae genomic clone AG-ND-127D17	BH403595	786	296.2	AG-ND-127D17, AF-ND-TAM Anopheles gambiae genomic clone AG-ND-127D17
2	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
3	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
4	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
5	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
6	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
7	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
8	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
9	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
10	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
11	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
12	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
13	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
14	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
15	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
16	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
17	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
18	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
19	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
20	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
21	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
22	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
23	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
24	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
25	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
26	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
27	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
28	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.
29	BH403595	786	269.1	genomic survey sequence.	BH403595	786	269.1	genomic survey sequence.

REFERENCE 1 (bases 1 to 786)  
Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Slim, C., Loftus, B.J., Ren, C., Huff, E.R., Carille, J.L., Black, K., Zhang, H.-B., Gardner, M.J., and Collins, F.H.  
Anopheles gambiae: Hexapoda: Insecta: Pterygota: Neoptera: Endopterygota: Diptera; Nematocera; Culicoidae, Anopheles.

Construction of a BAC library and generation of BAC end sequences from a genomic library of the African malaria mosquito Anopheles gambiae  
Mol. Genet. Genomics 268 (6), 720-728 (2003)

Other\_GSSs: AG-ND-127D17.TR  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Ref: 301 838 3543

Email: bjoftus@etig.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the midgut. The material was derived from mixed sexes of larvae. The BAC library was constructed at the University of Texas at Austin, Biology Department, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 For

Class: BAC ends.

Location/Qualifiers

1. .786  
/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-127D17"  
/clone\_lib="ND-TAM"  
/note="Vector: pGBAC1; Site\_1: HindIII"

BASE COUNT 183 a 211 c 148 g 244 t

Query Match 13.2%; Score 296.2; DB 28; Length 786;

Best Local Similarity 66.5%; Pred. No. 1.2e-75;

Matches 490; Conservative 0; Mismatches 233; Indels 14; Gaps 4;

QY 505 CTCAGGAAACAGCTCGCAACCTATGGCTTTAAACTTTGGT-TTTCAGGTGG 563

DB 728 CTACCGGTACGTTGCTTTGGANTCTATGGGATTACAAACCAAGGTATTTGGCGGTGC 669

QY 564 CAGACAGATGTATGGAGCCTCGAAGAGATGTATCTGGGGA----GCAGAAACCGAAT 619

DB 668 CCGTACAGATGTATGAAACCGCATAGATGTATCTTGGGAAGTAGGAACACTGGC 609

QY 620 GCGTGGGACAGAGCGTATGAGGTGACGAGAGCTCGGAAATCGCGGAGCCGTAC 679

DB 608 TGGGAACAGATAAAGATATACAGGGGATCGTCTGGAGATCCGCTGGCAGCAACA 549

QY 680 AATGGGACTCATCTATTAACCCCGGAAGACCAACGCAAGCCAGCCATCGCTG 739

DB 548 CAAATGGGGCTTCTATGTAAATCTGAAGGACCGGAAGAACCGATCGCGTGGCAG 489

QY 740 CTCGCCCTCATATTCGACATCTTGGCCGATGCGCAATGAATGACGAAGAACCCTGG 799

DB 488 CTCCTAAAGATATCCGTATACATTTTGGCCGTATGGCATGAATGACGAGAACTGTAG 429

QY 800 CTCCTACACGGGTGGACACCTCTCGAAACCAATCGGTGCTCGATGCGGGAAT 859

DB 428 CCGTAAATTCGTGGTGGCCAGCCATCTGGTAAATCTATGGTAAATCTATGGTAAATCT 370

QY 826 ATTCGCCGCGAGAGCTGCCGCCGACGATTTGGAAGAAATGACGCTGGGTGGGAACA 919

DB 369 --GTAGGAGCTGAGCTGAAGGATCCGGAATTTGAAGCTCAGGNTATAGTTGGGCAAGTA 312

QY 920 CTTACGGCACCGACAGCGGTGGGATACATCATCAGCTGAGAGCGCTCGAGCA 979

DB 311 AATTGGGACAGCTCCGAGCCAGCTGTATACCACTGGTCTGGAGTATCTTGGACCA 252

QY 980 AGACCCCTACTCAATGAGGCAATCACTTTTGAAGAACTCTTGGTATGAGGTGGAGC 1039

DB 251 AATACACCAACCAATGAGCAACAGCTCTTCAGCACTTATTCAGTATGAATGGAGC 192

QY 1040 TTACCAACATCCAGTGGAGGTATCATAGTGGAAACCAAAAGAGCGGTGCCGGGCTGGCA 1099

DB 191 TACGGAAGACCGCTGCGGTGCTCTCATGTGATTTGCNAAAGA-----TGGCCGAGATA 138

QY 1100 CAAATCCGGATGACATGATCCAGCAAGTGGCAGCTCCATTTATGCTCACTACGAGCA 1159

137 TTATTCGAATGCTTTTGATGCAGCAAAAAAGGTCCGCTTATATGCTTTACAACCTGACC 78

QY 1160 TGGCGTGGCACTGACCTCGTATTCAGAAAATTTCTGCAGCGTACTATGAAACCCGTG 1219

DB 77 TTTCTATGAATGTTGACTTATTTATGAGAAATATCAGAGCTTTCTTGATGAAATCCGG 18

QY 1220 ATGAGTTTTCGATGCT 1236

DB 17 AAGAACTGGCAGAGCT 1

RESULT 2

LOCUS AF075865

DEFINITION AF075865 Salmonella typhimurium LT2, Lambda DASH II Salmonella typhimurium genomic clone 1079-T7, genomic survey sequence.

ACCESSION AF075865

VERSION AF075865.1 GI:3320735

KEYWORDS Salmonella typhimurium

SOURCE Salmonella typhimurium

ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.

REFERENCE Wong R.M.-Y., Wong, K.K., Benson, N.R. and McLielland, M. Sample sequencing of a Salmonella typhimurium LT2 lambda library: the effect of the choice of restriction enzyme on the sequencing results. FEMS Microbiol. Lett. 173 (2), 411-423 (1999)

TITLE JOURNAL

PMID 99243757

PUBMED 10227170

COMMENT Contact: McLielland M  
Molecular Biology  
Sidney Kimmel Cancer Center  
University of California, San Diego  
Email: mclielland@lifsci.ucsd.edu

Class: Shotgun. Location/Qualifiers

1. .1102  
/organism="Salmonella typhimurium"  
/mol\_type="genomic DNA"  
/strain="T7"  
/cloned\_into="pTZ19"  
/cloned\_into="1079-T7"  
/clone\_lib="Salmonella typhimurium LT2, Lambda DASH II"  
/note="Vector: Lambda DASH II; sequenced using Li-Cor sequence"

BASE COUNT 235 a 278 c 334 g 251 t 4 others

ORIGIN

Query Match 12.6%; Score 283; DB 28; Length 1102;  
Best Local Similarity 58.6%; Pred. No. 1.2e-71;  
Matches 595; Conservative 0; Mismatches 389; Indels 31; Gaps 5;

QY 1225 TTTTCAGATGCTTTCCGAAGCATGGTCAAACTGACACAGAGATATGGGACCAAG 1284

DB 17 TTTTCAGATGCTTTCCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 76

QY 1285 GTCCCTACTCTGGGACAGATGCTCTGCAAGAGACTCATCTGCGAGAGCTTACCA 1344

DB 77 GCGCGTTACATCTGGCAGCGAATGCGCAAGAGATGATCTGCGAGAGACCGTTCGCG 136

QY 1345 GATGTAGACATCTCTTGTAGACGAAAACATATTTGAAGGCTTAAAGCCAAATCTGT 1404

DB 137 CACCGCTCTATTCGCGACGCGGAGGAGCATATTCAA---CTGAAACGCGGCGTCTGT 193

QY 1405 GAATCGGAGTACGCTGAACGAGCTGTGAAGCAGCGGATGCTGTGCTGCTACTTTT 1464

DB 194 GCATCCGGGCTTCTATTATGAGAGATGTCTTTCGGTTGCTTGGGATCCGCGCTCTACTTTC 253

QY 1465 AGAACTCTGACAGCGGGCGGTGGCAAGCGGTGCGAGCTATAGCTGCGCCCAACAAAA 1524

DB 254 CGCGGCGGATAGCGGTGGCGGCGGTACACGCGGCGGCGGTCTGGCATTAAGCGGCTCAGCGC 313





FEATURES		source	
Best Local Similarity	64.68; Pred. No. 6.6e-55;		
Matches	387; Conservative	0; Mismatches	212; Indels
0; Gaps			
Qy	90	TGCGAGTGGCGGCGCAAAACAGGAGTGTGGTGGCCACACATGCTCAACCTCGGCACTT	149
Db	82	TGGCGGTGGTGGTACCGCCACACAGGAGTGGTGGCCAAATAGCTGACACCAAGATCTT	141
Qy	150	AGCCACACATTCATGCTATGAGCCCAACAGCCGGATTTGACTATGCCGAGAGTT	201
Db	142	GGCCACACATTCATGCTATGAGCCCAACAGCCGGATTTGACTATGCCGAGAGTT	201
Qy	210	TAAAGAGCTAGCTATGCGAGCGGTAAAGAGAGCTGGCAGCGCTAATGACAGATCA	269
Db	202	CAGAAGCTTGGATTAATAGTGCCTCAAGAGAGCTTGAAGAGCTTCATGACCGACT	261
Qy	270	GGAGCTGGCGGCGCAAGATACAGCTATGATGGCCCTGCTTTATAGCATGGGTGGCA	329
Db	262	GGAGCTTGGCGGCGCAAGATACAGCTATGATGGCCCTGCTTTATAGCATGGGTGGCA	321
Qy	330	CAGGCGCGGCACTACGCTATGCTGATGGTATGGCGGTGGTGGCGCTCGGCTCACAGCG	389
Db	322	CAGGCGCTGATACCTACGCTATGCTGATGGTATGGCGGTGGTGGCGCTCGGCTCACAGCG	381
Qy	390	CTTGCGGCTGATACCTACGCTATGCTGATGGTATGGCGGTGGTGGCGCTCGGCTCACAGCG	449
Db	382	ATTGCTGCTCTCACTCACTACGCTGATGGTATGGCGGTGGTGGCGCTCGGCTCACAGCG	441
Qy	450	TGCGCCATCAAAACAAATACGCTGCAAAATCTCTGGCGGATCTAATGATATCTAC	509
Db	442	TGCGCCATCAAAACAAATACGCTGCAAAATCTCTGGCGGATCTAATGATATCTAC	501
Qy	510	AGCAAGCTACCTTGGCAATCTGCGGCTTAAACCTTTGATTTGCTGATGGTGGCGAGC	569
Db	502	TGCAATCTGCGGCTTGGCAATCTGCGGCTTAAACCTTTGATTTGCTGATGGTGGCGAGC	561
Qy	570	AGATGATATGGGAGCTGAGAGAGATCTATATCTGGGAGCAGCAACCGAATGGCTGGGAGA	629
Db	562	CGACATCTGGGAGCTGAGAGAGATCTATATCTGGGAGCAGCAACCGAATGGCTGGGAGA	621
Qy	630	CAGGCTGATGAGCTGAGGAGCTGCAAAATCTGCGGAGCGGAGCGCTCAATATGGGAC	688
Db	622	CAGGCTGATGAGCTGAGGAGCTGCAAAATCTGCGGAGCGGAGCGCTCAATATGGGAC	680
RESULT 6			
BO752115	762 bp	nrna	linear
EST343678	DSCT	Colletotrichum trifolii	cDNA clone pBSCT10-12, nrna
DEFINITION	sequence.		
ACCESSION	BO752115.1	gi:21907520	
VERSION	BO752115		
KEYWORDS	EST.		
ORGANISM	Colletotrichum trifolii		
REFERENCE	Colletotrichum trifolii		
AUTHORS	Smaad,D.A., Dickman,M., Town,C.D., Van Aken,S., Otterback,T.,		
TITLE	ESTs from mycelia of Colletotrichum trifolii race 1		
JOURNAL	Unpublished		
COMMENT	Contact: Deborah A. Smaad		
	Department of Plant Pathology		
	University of Minnesota		
	495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA		
	Tel: 612 625 1243		
	Fax: 651 649 5058		
	Email: debby@puccini.crl.umn.edu		
	Accession name: M58J212V More information is available at:		
	www.ncbi.nlm.nih.gov		
	Seq primer: 191A ATA CGA CTC ACT ATA 959 C).		

B0751801 707 bp mRNA linear EST 18-JUL-2002  
 LOCUS EST632364 DSCC Colletotrichum trifolii cDNA clone pBSC77-68, mRNA  
 DEFINITION sequence.  
 ACCESSION B0751801  
 VERSION B0751801.1 GI:21907206  
 KEYWORDS Colletotrichum trifolii  
 ORGANISM Colletotrichum trifolii  
 EST. 18-JUL-2002  
 SOURCE Colletotrichum trifolii  
 ORGANISM Colletotrichum trifolii  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;  
 mitospicric Phyllachoraceae; Colletotrichum.  
 1 (bases 1 to 707)  
 REFERENCE Samad,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,  
 Cheung,F. and Fraser,C.M.  
 1999. Isolation and characterization of Colletotrichum trifolii race 1  
 isolates from mycelia of Colletotrichum trifolii race 1  
 unpublished.  
 JOURNAL Contact: Deborah A. Samad  
 COMMENT Department of Plant Pathology  
 University of Minnesota  
 495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA  
 Tel: 612 625 1243  
 Fax: 651 649 5058  
 E-mail: dsamad@dep.umn.edu  
 TIGR sequence name: MT5663TV More information is available at:  
 www.medicalgo.org  
 Seq primer: (5'-ATA CGA CTC ACT ATA ggg C).  
 Location/Qualifiers  
 1: 707  
 /organism="Colletotrichum trifolii"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:5466"  
 /clone="pBSC77-68"  
 /tissue\_type="mycelia"  
 /dev\_stage="Young, actively growing mycelia (3 days after  
 inoculation), grown in liquid culture (cutin minimal medium  
 containing 0.1% glucose)."  
 /clone\_lib="DSCC"  
 /note="Vector: pBluescript SK+; Site\_1: EcoRI; Site\_2:  
 EcoRI; isolate: 2sp2; cDNA was prepared from polyA+  
 enriched RNA. The cDNA was ligated into Lambda gtl1 from  
 Stratagene and packaged using GigaPack packaging extracts.  
 An aliquot of the amplified library was used to transduce  
 E. coli and the transductants were purified from a liquid  
 lysate. The cDNA library was screened with a 300bp  
 digestion and ligated into phuscript SK+. Aliquots of  
 the ligation were used to transform E. coli DH5alpha which  
 were plated onto medium with X-gal for selection of  
 recombinants."

# FEATURES

source  
 1: 707  
 Location/Qualifiers  
 11..48: Score 254.8; DB 13; Length 707;  
 Best Local Similarity 56.24; Pred. No. 1.9e-63;  
 Matches 367; Conservative  
 0; Gaps 0;

QY 79 CTTAAGCAAGATGCAAGTGGCGCCACCAAAACAGGAGATTGGTGGCCCAACATGCTCAAC 138  
 DB 140 CTGATGATCAACAGCGGGGGGGGCGGCGATCATCAAGACGCGATGTTGGCGGAGATCGCTGAG 199  
 QY 139 CTGATGATCAACAGCGGGGGGGGCGGCGATCATCAAGACGCGATGTTGGCGGAGATCGCTGAG 198  
 DB 200 CTAACATCTCCCGCCAGCAGACGCCGCTGACCAACCGCGTGGCCAGGACTTTGACTAC 259  
 QY 199 CCGAAGAGATTGAAGACATAGATCTGGCAGCGGTTTAAAGAGACCTGGCAGCGCTAATG 258  
 DB 260 GTCCGCGGCTTCAGTCGCTGAGTACGAGGAGGCTCAAGAGAGACCTCAGCGCCTGATG 319  
 QY 259 ACAGATTCACAGACTGTGGCGAGCAAGATTACGTCATATGTCCTTCTTTTACCG 318  
 DB 320 ACCGACTCCAGAGTGTGGCTCCGCACTTTTGGCCACTACGCGGCTGTCTTATCCG 379

QY 319 ATGGCTGTGCACACAGCGCCGCACTACCGATCGGTGATGGCTGTGGCGGTGGCTCC 378  
 DB 380 ATGGCTGTGCACACAGCGCCGCACTACCGATCGGTGATGGCTGTGGCGGTGGCTCC 439  
 QY 379 GGCTCAACAGCGCTTGGCGGCTCTCAATAGTGGCGAGCAATGCGCATCTGGATAACGA 438  
 DB 440 GGCAACAGCGCTTGGCGGCTCTCAACAGTGGCGGCAATGTGACCTCGACANAGGCC 499  
 QY 439 CCGTCTGCTTGTGGCCCATCAACAAATACGGTTCGAAATATCTCTGGGGCGATCTA 498  
 DB 500 CCGTCTGCTGCTGGCCCATCAACAAATACGGTTCGAAATATCTCTGGGGCGATCTA 559  
 QY 499 ATGATATCAACAGAAACCTAGCTTGGAAATATGCGCTTAAACATTTTGGTTTGA 558  
 DB 560 ATGATCTGGCGGCAACAGTGGCGCTCGAGTCCATGGGTTCACAGAGCGCGCTTCTCC 619  
 QY 559 GTGGCAGACAGATGTGGAGCGCTGAGAGATGTATCTATCTGGGAGCAGAAACGAA 618  
 DB 620 GGAGCGCTGCGACACTGGGAGCGCAGATGCTCTATGTGGGCGGAGAACAC 679  
 QY 619 TGGCTGGGAGACAA 632  
 DB 680 TGGCTGGGAGACAA 693  
 RESULT 8  
 B065388 710 bp mRNA linear EST 26-AUG-2002  
 LOCUS Fgr\_7.G22\_T3 Nitrogen-starved mycelia Gibberella zeae cDNA, mRNA  
 DEFINITION sequence.  
 ACCESSION B065388  
 VERSION B065388.1 GI:22505677  
 KEYWORDS Gibberella zeae  
 SOURCE Gibberella zeae  
 ORGANISM Gibberella zeae  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
 REFERENCE Trail,F., Wu,R., San Miguel,P., Hølgren,R.G. and Kistler,H.C.  
 1998. A set of sequence tags from Gibberella zeae (anamorphic  
 Fusarium graminearum)  
 Fungal Genet. Biol. 38 (2), 187-197 (2003)  
 JOURNAL  
 MEDLINE 22508120  
 PUBMED 1250255  
 COMMENT Contact: Frances Trail  
 Department of Plant Biology  
 Kansas State University  
 East Lake Hall, 14824, USA  
 Tel: 517 433 2939  
 Fax: 517 353 1926  
 Email: trail@ksu.edu  
 Plate: 7 row: G column: 22.  
 Location/Qualifiers  
 1..710  
 /organism="Gibberella zeae"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:5518"  
 /clone\_lib="Nitrogen-starved mycelia"  
 /note="Vector: Uni-ZapII; Site\_1: EcoRI; Site\_2: XhoI"  
 BASE COUNT 139 a 238 c 192 g 141 t  
 ORIGIN  
 Query Match 11.18; Score 249.4; DB 13; Length 710;  
 Best Local Similarity 60.88; Pred. No. 7.6e-62;  
 Matches 425; Conservative  
 0; Gaps 1;  
 QY 1480 CGCGCGCGTGCACAGCGTGCACATATACGACTGGCGCCCAACAAAAGACTGGGAGTAAC 1539  
 DB 12 COTGGTGGTGCACAGCGTGCCT 71  
 QY 1340 NACCCTCAGCACTTGGCCAGGCTACTCAAACTAGAGAGTATCCAGAGAGACTTTCAC 1599

```

Db 72 ACCCTCTCTGCTGGCCAGGTCTCTACAGGCTTTTGAGAGATCCGAGAGACTTTCAC 131
Qy 1600 CAGGCGCAATCAGATATACAGAGCAGTATCTGTTGGCCGACTGATTTGCTGGCGGCTGT 1659
Db 132 GACTCTCTAGTCCACGGCAGGCGCATCTCTCTTGTGAGCTGATCGTCTTGTCTGTCT 191
Qy 1660 GCGGCTGTAGAMAGCTGTGAMAGATGCTGGCATGAGGTGGCTTTCACCCG 1719
Db 192 GTGCTGCTGAGAGAGCGCCAGAGAGGCTGATTAACGTACACGCTGCCCTTCTACTCT 251
Qy 1720 GGAAGCAGCGATGCCACCGCTGAGCAGAACGTGTGAAGCTTTTGAAGCACTAGAGCA 1779
Db 252 GCGCGTACGATGACACCGCAGCAGCAGCAGCTGCTGCTGCTCACTTCGAGCC 311
Qy 1780 GCGGCTGACGCGCTGTAGAACTACATTAACCGGAGCATTAAGTATCCCTGTAGAAATG 1839
Db 312 GCCGCCAGCGTTTCGAAATATGCGCGCTTCCACAGCCGTCTCAAGCTCGAGCAGTG 371
Qy 1840 CTCGTGAGCAGCGGCGAGCTTCTGTGCTTTTGGCAGCAGAAATGATCTTTGTAGCG 1899
Db 372 CTCATAGCGTCCGCGAGCTCTCACTCTACGCTCCCGGAGATCACCGCTCTGTGGGT 431
Qy 1900 GGTATGAGCTATGGGCAACGATTCAGCTGCTTCCGAGGTGGGATGTTTACAAATAG 1959
Db 432 GGTCTGTGGTGCCTCAACGCACTTGGGAGCGATCATCCACGATGTTCTCCACAGCGC 491
Qy 1960 CCGGCTGAGCTATCCATGACTCTGTTGTTAACTCTGAGACTCAACACTAAATGGCGA 2019
Db 492 CCTGGTGTCTGTCCACAGCACTTCTGCTCAACTCTCGACATCAGACCGAGTGAAG 551
Qy 2020 GCGAGGATGATGACGAAGTTTTCAGGCGAGAGCTTCAAACTGGGCGAGTANAG 2079
Db 552 CCGCTGGCGA---CGCGGACTGTTTCAGGGGTATTGACCGCAGACGCGCAACAAGAG 608
Qy 2080 TGGAGTGGCAGCGGAGTGAAGCTGATCTTCGATCCCAATTCGAGCTTAAGAGCCCTCGCA 2139
Db 609 TGGAGTGGCAGCGGAGTGAAGCTGATCTTCGATCCCAATTCGAGCTTAAGAGCCCTCGCA 2139
Qy 2140 GAAGTGTACGCGTGTGCAGATCTGCTGCTTGGATCAGCTGAGCTCCGCGCGAGCT 2178
Db 669 GAGCATATGCTGAAGCTGGCGGTAGCAGAGCTGTT-707

RESULT 9
LOCUS AF075904
DEFINITION Escherichia coli B173-73, genomic survey sequence.
ACCESSION AF075904
VERSION AF075904.1 GI:3320774
KEYWORDS GSS.
ORGANISM Salmonella typhimurium
Salmonella typhimurium
Bacteria; Proteobacteria; Enterobacteriaceae; Salmonella.
REFERENCE Wong, R.M.-Y., Wong, K.K., Benson, N.R., and McLielland, M.
AUTHORS Sample sequencing of a salmonella typhimurium L72 lambda library:
TITLE comparison to the Escherichia coli K12 genome
JOURNAL FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
PUBMED 10227170
COMMENT Contact: McLielland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclielland@lifsc.sdsu.edu
Class: shotgun.
Location/Qualifiers
/mol_type="genomic DNA"
FEATURES
source

```

```

/strain="L72"
/Accession="AF075904"
/CloneLib="Salmonella typhimurium L72, Lambda DASH II"
/Note="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT 188 a 248 c 283 g 199 t
ORIGIN
Query Match 10.98; Score 244.6; DB 28; Length 918;
Best Local Similarity 58.9%; Pred No 2,3e-60;
Matches 532; Conservative 0; Mismatches 334; Indels 37; Gaps 5;
Qy 900 GAGCTTGGSGTGGAAAAACACTTACAGCAGCCGAGACGCTGGGATACATCACACGATGG 959
Db 24 GGGCTTAGTGTGGGCCAGCAGCTATGTATGTGGCTGGCCGAGATGATCACTTCCCG 83
Qy 960 ACTTAGAGCGCCTTGACACAGACCCCTTACTCAATGAGCAGATACCTTTTGGAACTCT 1019
Db 84 GCTGGAAGTGTCTTGACAGCAGACGCGACCCAGTGGAGCACTATTCTTCGGAACCT 143
1020 CTTTGTGTACAGTGGGAGCTTTACAAAAGTCCAGCTGGAGCTTATCATCGGAACCAA 1079
Qy 1476 GTTATGATGCTCATTACAGGACCTGGCGCTGGCAGTGGACCTGATTACGAAAAATTTCTCG 1199
Db 258 AACCTGTGCTCCGAGCTGAGCTGGCTTTGATCGGAGTTCGAGAGATTTTCGCC 317
Qy 1200 ACGGTACTACAAACCTGTAGTGTTCGAGATGCTTCCCGAAGCATGTTACAACT 1259
Db 318 TCGTTTCTTAAGCATCC--GCAGCCTTTTAATGAAGCTTTGCTCGTGGCTTCAACT 376
1260 GACACAGAGATATGGGACCAAAAGTGGCTTACCTGGGACGAGAGTGGTGAAGCAC 1439
Qy 497 CAA--CTGAAAGAGCGCTGCTGCTATCGGCGCTTCTATTGACAGATGGTTTCGGT 553
Db 1440 GGCATGGGCTCTTGATCTACTTTTAAAGAACTTGACAGAGCGGCGGTGGCACAGTGC 1499
Qy 554 TCCCTGGGATCCGCGTCTACTTTCCGCGCGCGGATTAAGCTGGCGGCTTACCGGCG 613
Db 1500 AGTATACAGTGGCGCCCAAAAGATGGGAGTAAACAAACCTTCAGCACTTTCGCAG 1559
Qy 614 GGTGTGGCATCTGAGCGCTGAGCGCTTGGATGTCACCCGTTGCG-----GCTCG 667
Qy 1560 GGTACTCAAAACATAGAGGTATTCAGGAGGACTTTAACCGAGCGCAATCAGATACAA 1619
Db 668 CGTCTCGCGGTATTAGAGAGATCCAGAA-----AACGACGAA 706
Qy 1620 AGCAGTATCGTGGCGGACCTGATGTGCTGGCGGCTGCTGGGTGTAGAAAAGCTGC 1679
Db 707 TAAACCTCTGCTGGCGATTTATTTCTGCGGCGGCTGCTGGGTATGAGAGCGGCG 766
Qy 1680 AAAAGATGCTGGCATGAGGTGCGAGTGCCTTTCAACCGGAGCAGCGATGTCACCGC 1739
Db 767 CGCTGCTGGCGGTGTGAGCATCAGCGTACCTTTTTCGCGCGCGCGGCTGGATCGCGGTCA 826
Qy 1740 TGAGCAACCGATGTGAGAGCTTTTCAGAGCACTAGACCCAGCCGCTGACGCGTTTAAAG 1799
Db 827 GGTCAAGCGCATCGAGATGTTCTGCTGCTGTGACCGCATTTGCTGATGGATTCGTTA 886

```







•

```

ORIGIN
Query Match      10 44; Score 231.2; DB 13; Length 639;
Best Local Similarity 67.04; Pred. No. 4.3e-57;
Matches 331; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 79 CTTAAGCAAAATGTCAGTGGCGGCAACAAACAGGAGTTGTGGCCCAACATGCTCAAC 138
DB 144 CTGATGACACAGCGCGGCGGCGGCTCATGAGACCGGCTGTGGCGGATGCGCTGAG 203
QY 139 CTGGGCTATTATACGCAACATCATGCTATGCGACCAACAGCCGGAATTTGACTAT 198
DB 204 CTCAACATCTCCGACGACAGCGCGGCAACACGCGTGGCGAGGACTTTGACTAC 263
QY 199 GCGAGAGGTTTAAAGAGCTAGATCTGCGACGCGTTTAAAGAGACCTGCGACGCTATG 258
DB 264 GTGCGCGCTTGTGCTGCTGACTACGAGGGGTGCGAGAGACTTCAGCGCCCTG 323
QY 259 ACAGATTACAGAGCTGTGGCGGCGGCTATAGTGATATATGAGCCCTCTTTATACG 318
DB 324 ACGACTCTCCAGGACTGTGGGCTGCGGACTTTGGCCACTACGAGCGGTGTGTCATCCG 383
QY 319 ATGGGTGGACAGCGCGGCGGCTACCGTATCGGTGATGGCGGTGAGTGGCGGTGCTCC 378
DB 384 ATGGCGTGGACAGCGCGGCGGCTGCTCCGCTTACAGCGGAGCGAGGTGGTGAG 443
QY 379 GCGTCTACAGCGGCTGCGCGCTCTCAATAGCTGCGCAGACATGCAATCTGGATTAAGCA 438
DB 444 GCGCAGCAGCGGCTGCGACGCGCTCAACAGCTGCGCGGACATGTGACGCTCGACAGCGC 503
QY 439 CCGTGTCTGTGTCGCGCAATCAACAAATAAGTACGCTGMAAATCTCTCTGGCGGATCTA 498
DB 504 GGTGCGCTGTGTCGCGCAATCAACAAATAAGTACGCTGMAAATCTCTCTGGCGGATCTA 563
QY 499 ATGATACTACAGAAAGTGTGCTGGAAGTACTGGGCTTTAAAGCTTTTGGTTTGA 558
DB 564 ATGATCTGCGCGCGCAAGTGGCGCTCGAGTCCATGAGTCCATGGTTCCAGACGCGGCTCTCC 623
QY 559 GTTGGCAGACGA 572
DB 624 GGAGGCGTCCCGA 637

RESULT 15
B0752114/c 633 bp mRNA linear EST 18-JUL-2002
DEFINITION B0752114 DSCT Colletotrichum trifolii cDNA clone pDSCT10-12, mRNA
SEQUENCE
ACCESSION B0752114
VERSION B0752114.1 GI:21907519
KEYWORDS EST.
SOURCE Colletotrichum trifolii
ORGANISM Colletotrichum trifolii
REFERENCE B0752114 DSCT Colletotrichum trifolii cDNA clone pDSCT10-12, mRNA
AUTHORS Sakuraya, Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Nectria; Nectriaceae; Nectria; Nectriaceae; Phyllachoraceae;
1 (bases 1 to 633).
Cheung, F. and Fraser, C.M.
ESTs from mycelia of Colletotrichum trifolii race 1
Other-ESTs: EST632678
Contact: Deborah A. Samac
Department of Plant Pathology
University of Minnesota
495 Borling Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debbys@umn.edu
TIGR sequence name: WTSJ121K More information is available at:
www.medicalgen.org
Seq primer: 5'-GATC (CTA GAA GTA GAT CC).
Location/Qualifiers

```

```

source
1. .633
/organism="Colletotrichum trifolii"
/strain="Race 1"
/db xref="taxon:5466"
/clone="pDSCT10-12"
/tissue_type="mycelia"
/dev_stage="Young, actively growing mycelia (3 days after inoculation) grown in liquid culture (cutin minimal medium with 1% glucose)."
/lab_building="PDS1000A"
/clone_lib="DSCT"
/note="Vector: pBluescript SK+; Site:1: EORI; Site:2: EORI; isolate: 2p2; cDNA was prepared from polyA+ enriched RNA. The cDNA was ligated into Lambda gt10 from Stratagene and packaged using GigaPack packaging extracts. An aliquot of the amplified library was used to transduce E. coli. The cDNA inserts were purified from the liquid lysate. The cDNA inserts were gel purified after PCR digestion and ligated into plasmid SK+. Aliquots of the ligation were used to transform E. coli DH5alpha which were plated onto medium with X-gal for selection of recombinants."
BASE COUNT 102 a 235 c 186 g 110 t
ORIGIN
Query Match      10 38; Score 231.2; DB 13; Length 633;
Best Local Similarity 63.18; Pred. No. 1.7e-56;
Matches 391; Conservative 0; Mismatches 223; Indels 6; Gaps 2;

QY 1618 AAGCAGTATCTGTGGCGACCTGATGTGCTGGCGGCTGTGGCGGCTGTAGAAAAAGCT 1677
DB 628 AANAAGTTCTTCCGCGACGCTATGCTGTGGCGGCTGTGGCGGCTGTAGAGCGCG 569
QY 1678 GCAAAAGATCTGCGCAAT--GAGTGTGACGTGCTTTCAACCGCGGACGAGCGATGGC 1734
DB 568 GCCAGGAACCGCGGCGACACGACGCTCAACGCTTCCATCACGCCGCGCGCGACGCC 509
QY 1735 ACGCTGTAGCAACGATGTGGAGCTTTTCAGACACCTAGACGCGGCTGACGGCTTT 1794
DB 508 TGGCAGAGACGACGAGTGTGCTGCGTGTGACACCTCCAGCTTTGGCGAGGGTTC 449
QY 1795 AGAAACTACATTAACCGGAGCATAAATATCGCTGAGGAATGCTGTAGACCGGGG 1854
DB 448 GCAACTACGAAAGATCGAACGAGCGGCTCAAGACGAGAGTTCCTGTGACCGCGCG 389
QY 1855 CAGCTGTGCTGTGGCGACCAAGATGCTGTTGTGAGCGGATATCCGTGTACTG 1914
DB 388 CAGCTCTCTACCTCTGCGCGCGGACTACGAGTCTCTGTGGTGTCCGCTGCTG 329
QY 1915 GGCACCAACTACGAGCGTGTGCGACGATGAGTGTTCACAAATAGCGGGGTGACGTATG 1974
DB 328 GGGGCCCACTACGAGCGTTCGCGGCGGCGGCGCTTTCACAGAGCGCGGCTGGCTGAGC 269
QY 1975 ATGACTTCTTTTAAACTCTCTAGACCTCAACACTAAATAGCGAGCGAGCATATCA 2034
DB 268 ATGCTTTCTTTGATCTCTGCGCTGCGACGCTGCGTGTGCGATTCGAGGCTGTATGCCAG 209
QY 2035 GACAAAGTTTGAAGCGAGAGACTTCAAACTGGCGGAAGTAAAGTGAAGTGCACCGCG 2094
DB 208 GAC--GTGTACGAGCAGACCGCGGCGGCGGCGCTTTCACAGAGCGCGGCTGGCTGAGC 152
QY 2095 GTAGCTGATCTTCGATTCGAAATCCGAGCTTAGAGCCCTCGAGAGTGTGTAGCGTGT 2154
DB 151 GTGACCTGTGCTGGGTGCGGCGGCGGCGGCTGCGTGTGCGATTCGAGGCTGTATGCCAG 92
QY 2155 GCAGATTCGAGAAAGTTTGTAAAGATTTTGAAGGCTGCGCAAGGTAAATGATGAC 2214
DB 91 GCGGATGCGGCGGCGAGAGTTTGTGAGGAGACTTTGTGAGCGGCTGGGATGAGGTGATGAC 32
QY 2215 CTGACCGCGGTGTGATCTGAA 2234
DB 31 CTGATCGGTTCCACCTGAA 12

```

Search completed: October 8, 2003, 15:29:20  
Job time : 4640.63 secs

GenCats version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:31:22 ; Search time 60.1769 Seconds  
(without alignments)  
1985.061 Million cell updates/sec

Title: US-09-884-889-8

Perfect score: 4002  
Sequence: 1 MENVHSGSYTWTGK.....VKDFVNAKVMDFDLK 745

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_13Jun03.\*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4002	100.0	745	19_AAW33810	Microscilla furvescens.
2	2544	63.6	740	18_AAW26595	Microbacterium tube
3	2544	63.6	740	19_AAW31343	Microbacterium tube
4	2544	63.6	740	20_AAW95398	M. tuberculosis ca
5	2541	63.5	740	20_AAW95399	M. tuberculosis ca
6	2290.5	57.2	753	19_AAW33809	Alcaligenes (Deley
7	2290.5	57.2	753	23_AWG31040	Alcaligenes catala
8	2178.5	56.9	740	18_AAW7870	M. tuberculosis H37
9	2178.5	56.9	735	20_AAW78363	M. tuberculosis H37

	10	2201.5	55.0	723	23	ABW78212
	11	1695.5	42.5	731	11	AAW08205
	12	886	21.1	733	22	AAW08205
	13	886	21.1	733	22	AAW08205
	14	675	16.9	1550	22	ABG09297
	15	455	11.4	201	22	ABG08828
	16	278.5	7.0	279	23	ABW93261
	17	273.5	6.8	291	23	ABW05661
	18	266.5	6.7	246	23	ABW92257
	19	246	6.1	257	21	AAW26582
	20	246	6.1	257	21	AAW26582
	21	244	6.1	250	21	AAW08702
	22	244	6.1	250	21	AAW16805
	23	244	6.1	250	21	AAW26453
	24	244	6.1	250	21	AAW3523
	25	244	6.1	250	21	AAW45582
	26	244	6.1	250	21	AAW45585
	27	244	6.1	250	21	AAW45585
	28	244	6.1	250	21	AAW45585
	29	244	6.1	254	21	AAW3522
	30	240.5	6.0	287	21	AAW09460
	31	240.5	6.0	287	21	AAW3316
	32	240.5	6.0	287	23	ABW93244
	33	232	5.8	215	21	AAW09703
	34	232	5.8	215	21	AAW16806
	35	232	5.8	215	21	AAW26454
	36	232	5.8	215	21	AAW26454
	37	232	5.8	215	21	AAW45583
	38	232	5.8	215	21	AAW45586
	39	232	5.8	215	21	AAW45590
	40	232	5.8	354	21	AAW4544
	41	232	5.8	426	21	AAW4543
	42	232	5.8	426	23	ABW1634
	43	232	5.8	426	23	ABW1634
	44	225.5	5.7	287	21	AAW7832
	45	223.5	5.6	254	21	AAW09461

#### ALIGNMENTS

RESULT 1  
ID AAW33810 standard; Protein; 745 AA.  
XX AC AAW33810;  
XX DT 18-NOV-1998 (first entry)  
XX DE Microscilla furvescens catalase-53CAL.  
XX KW Catalase; epoxidation; hydroxylation; biosensor; paper bleaching;  
XX PW pasteurisation.  
XX OS Microscilla furvescens.  
XX SN W09800526-A1.  
XX PD 08-JAN-1998.  
XX PF 03-JUL-1997; 97WO-US16513.  
XX PR 03-JUL-1996; 96US-0674887.  
XX RE (RECO-) RECOMBINANT BIOTRANSLYSIS INC.  
XX AD Adhikary RS, Robertson DE, Sanyal I;  
XX WPI: 1998-086953/08.  
XX N-PSDB: AAW06555.  
XX New bacterial catalases, related nucleic acid vectors and transformed cells - used as oxidising agents and for detecting or

PT destroying hydrogen peroxide, e.g. in biosensors  
 PS Claim 1; Fig 2; 35pp; English.  
 XX  
 CC The present sequence is of the *Microscilla furvescens* catalase-53CAl.  
 CC Catalase-53CAl may be used to catalyze oxidation reactions such as  
 CC epoxidation or hydroxylation. The enzyme can also be used to detect or  
 CC destroy hydrogen peroxide, e.g. in connection with glyoxylic acid  
 CC peroxidation, colorimetry, conduct, immunoassay, pulp/paper bleaching and  
 CC detection of contaminants, etc.  
 CC Catalase-53CAl can be used to screen libraries for detection of the  
 CC catalase-53CAl gene.  
 CC purification of cells containing the enzyme. Fragments of the gene  
 CC encoding Catalase-53CAl can be used to identify related sequences.  
 XX  
 XX Sequence 745 AA:

Query Match 100.0%; Score 4002; DB 19; Length 745;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MENEKSSSTNTNTGKCFPTGGSQAGGCTKNRDPNMLGLIRHSSLSDPN 60  
 DB 1 MENEKSSSTNTNTGKCFPTGGSQAGGCTKNRDPNMLGLIRHSSLSDPN 60  
 QY 61 DPDPDYAEFKKDLAAVKDAAALMTDSQWPAQYCHGYPTIRMAHSACTVRIGDS 120  
 DB 61 DPDPDYAEFKKDLAAVKDAAALMTDSQWPAQYCHGYPTIRMAHSACTVRIGDS 120  
 QY 121 RGGGSSGSRFAPLNSWPNANLDRKLLMLPIKQYGRKISWADLMLTGNVALETMGF 180  
 DB 121 RGGGSSGSRFAPLNSWPNANLDRKLLMLPIKQYGRKISWADLMLTGNVALETMGF 180  
 QY 181 KTFGAGCRADYWEPEEDYVGAETEMLDGKRYEGDRELENLGAVOMGLITYNPEGSPG 240  
 DB 181 KTFGAGCRADYWEPEEDYVGAETEMLDGKRYEGDRELENLGAVOMGLITYNPEGSPG 240  
 QY 241 KPDPYAAARIDREFGRMANDEETVALTAGHGTGKTHGAADAERYGVREPAAGIEEM 300  
 DB 241 KPDPYAAARIDREFGRMANDEETVALTAGHGTGKTHGAADAERYGVREPAAGIEEM 300  
 QY 301 SLGWNNTYGTGHGADTITSLGEGATKPTQWNSNFFENLFGYEWELTKSPAGATQWPKP 360  
 DB 301 SLGWNNTYGTGHGADTITSLGEGATKPTQWNSNFFENLFGYEWELTKSPAGATQWPKP 360  
 QY 361 DQAGAGTIPDAHDPSKSHAFPLMTDLALRMDPDEKISRYRYENPDEFADAFANAYKL 420  
 DB 361 DQAGAGTIPDAHDPSKSHAFPLMTDLALRMDPDEKISRYRYENPDEFADAFANAYKL 420  
 QY 421 THRDGKNRYLGPVEPDELWDPDIPVSHPLVDENDLEGLKALLESGLITYSELVST 480  
 DB 421 THRDGKNRYLGPVEPDELWDPDIPVSHPLVDENDLEGLKALLESGLITYSELVST 480  
 QY 481 ANASASTFRNSDKGANGARITLAPKQWNVNPNQOLARVLTLEGTQEDFNQAQSDNRK 540  
 DB 481 ANASASTFRNSDKGANGARITLAPKQWNVNPNQOLARVLTLEGTQEDFNQAQSDNRK 540  
 QY 541 AVSLADLLVLACGAVEAKADGHEVQVYFNPGRADTAECTDVEALEPAADCFN 600  
 DB 541 AVSLADLLVLACGAVEAKADGHEVQVYFNPGRADTAECTDVEALEPAADCFN 600  
 QY 601 YIKPEKVSAREMVDRAQLLSLSPMTALVGMNRYLGTNDGSGRGVFTNRKQGLISND 660  
 DB 601 YIKPEKVSAREMVDRAQLLSLSPMTALVGMNRYLGTNDGSGRGVFTNRKQGLISND 660  
 QY 661 FYNVLDDLNTKWRASDESQVTEGRDFTKGVNSGTRVDLLTFCGNSLALAEVYGCAD 720  
 DB 661 FYNVLDDLNTKWRASDESQVTEGRDFTKGVNSGTRVDLLTFCGNSLALAEVYGCAD 720  
 QY 721 SEEFKPYDFYKAWAYMDLDRFDLX 745  
 DB 721 SEEFKPYDFYKAWAYMDLDRFDLX 745

RESULT 2  
 AAW26596  
 XX AAW26596 standard; Protein; 740 AA.  
 AC AAW26596;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 21-JAN-1998 (first entry)  
 XX  
 XX Mycobacterium tuberculosis wild-type catalase-peroxidase.  
 KW Isoniazid resistance; restriction fragment length polymorphism;  
 KW RFLP; katG gene; catalase; peroxidase; multiple drug resistance;  
 KW Isonicotinic acid hydrazide; INH; tuberculosis; antibiotic;  
 KW tuberculostatic.  
 XX  
 XX Mycobacterium tuberculosis.  
 PH Key Location/Qualifiers  
 FT Misc-difference 264  
 FT Misc-difference /note= "mutation to Thr confers INH resistance"  
 FT Misc-difference 315  
 FT Misc-difference /note= "mutation to Thr confers INH resistance"  
 FT Misc-difference 337  
 FT Misc-difference /note= "mutation to Cys confers INH resistance"  
 FT Misc-difference /note= "mutation to Leu confers INH resistance"  
 XX  
 XX US5658733-A.  
 PN  
 XX  
 XX 19-AUG-1997.  
 PD  
 XX  
 XX 07-APR-1995; 95US-0418782.  
 XX  
 XX 07-APR-1995; 95US-0418782.  
 PR  
 XX 18-APR-1994; 94US-0228562.  
 PR  
 XX (MAYO-) MAYO FOUNDATION.  
 PA  
 XX  
 XX Cockrell FR, Kline BC, Uhl JR;  
 XX WPI; 1997-424226/39.  
 XX N-PSDB; AAT90400.  
 DR  
 XX  
 XX Determining susceptibility of Mycobacterium tuberculosis strains to  
 PT Isoniazid - by detecting mutation(s) in the catalase-peroxidase gene,  
 PT katG  
 XX  
 XX Example 2; Column 31-36; 38pp; English.  
 XX  
 CC This polypeptide comprises a consensus of the Mycobacterium  
 CC tuberculosis wild-type catalase-peroxidase enzyme. Claimed  
 CC methods of rapidly identifying strains of M. tuberculosis which are  
 CC resistant to the tuberculostatic drug Isoniazid (INH) are based on  
 CC the discovery of 4 mutations in the wild-type katG gene sequence  
 CC (see AAT90400) which confer INH resistance and which coincidently  
 CC confer resistance to rifampin and streptomycin.  
 CC site. Primers (see AAT90401-02) are provided for the PCR  
 CC amplification of a katG gene from a test M. tuberculosis strain,  
 CC and further primers (see AAT90403-06) for use in the restriction  
 CC fragment length polymorphism analysis of the amplified gene and  
 CC hence determination of the susceptibility to INH of the strain.  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 XX  
 XX Sequence 740 AA;

Query Match 63.6%; Score 2544; DB 18; Length 740;  
 Best Local Similarity 64.3%; Pred. No. 1.4e-215;  
 Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

QY 2 ENUKHSGSSSTNTNTGKCFPTGGSQAGGCTKNRDPNMLGLIRHSSLSDPN 61  
 DB 3 EQHPPTTETTCGASNG-CPVGHKMYVEGGS -NQDWPFRNLKLVYGNPAPADPAG 59







XX 07-MAY-1997; 9705-0852219.  
XX (MAYO-) NATO FOUNDATION.  
XX Cockerill FR, Kline BC, Uhl JR.  
XX WPI; 1999-070099/06.  
XX N-PSDB; ANX0818.  
XX detection of Mycobacterium tuberculosis - by amplifying katG gene  
XX and detecting specific fragment, and optionally identifying  
XX INH-resistant strains by detecting specific mutation  
XX Claim 3; Page : 83pp; English.  
XX The invention relates to a novel method of detecting Mycobacterium  
XX tuberculosis. The method comprises amplifying the DNA in the samples to  
XX generate a detectable amount of amplified DNA comprising a catalase-  
XX peroxidase (katG) DNA fragment with sequence of bases 904-1523 of the  
XX M. tuberculosis strain H37Rv MC katG gene sequence. The method optionally  
XX further comprises determining if the katG DNA fragment has a serine to  
XX threonine mutation in codon 315 (S315R mutation), indicative of an  
XX isoniazidic acid hydrazide (INH)-resistant M. tuberculosis strain. The  
XX method uses a PCR reaction to amplify the DNA. This method is useful  
XX especially for human sputum useful for diagnosis. This method is useful  
XX as a major cause of human morbidity and mortality, and conclusive diagnosis  
XX and subsequent treatment depends on identification of the etiologic agent  
XX M. tuberculosis. INH has been used in tuberculosis treatment, but  
XX INH-resistant strains have emerged; the method allows such drug-resistant  
XX strains to be identified. The present sequence represents the amino acid  
XX sequence of the M. tuberculosis katG gene variant.  
XX The method is novel and useful for identification; it has been  
XX created by modifying the katG gene product given in Fig 7.  
XX Sequence 740 AA;

Query Match 63.5%; Score 2541; DB 20; Length 740;  
Best Local Similarity 64.2%; Pred. No. 2.6e-215;  
Matches 479; Conservative 89; Mismatches 168; Indels 10; Gaps 8;  
2 ENHIGSSSTVNTGKCPPTGSGLSKOSAGGKTNRDWPNNLGLIRBSSLSND 61  
3 EORPETITTTGAASNG-CPVVGKMKYVEGGG--NODMNPRLNKLQHPNADPMG 59  
62 PDVPAEKKLDAVKKDLAALMTDSQWMPADYGHGPFIRMAHMSAGTYRIGDGR 121  
60 AAFDAEVAIVDMLREIVETVSGWPAFGHGGPFIIRMAHMAHTRIDHGR 119  
122 GGGGSGSQFAPLNSWPNANLAKLLWPKYKGRKISWADLMLGNVALETMGFK 181  
120 GGAGGQNGRFPALNSWPNANLAKLLWPKYKGRKISWADLMLGNVALETMGFK 179  
182 TFGGGRADYVWPEEDYVNGATEWMLDKRYEGDRELENPGLAVQMLIYVNPESNGK 241  
180 TFGGGRADYVWPEEDYVNGATEWMLDKRYEGDRELENPGLAVQMLIYVNPESNGK 238  
242 PDPTAAARDIETFRGMAANDDEEVALTAGSTFGTKTGADAAKRYGPEAPAAQEDMS 301  
239 PDPTAAARDIETFRGMAANDDEEVALTAGSTFGTKTGADAAKRYGPEAPAAQEDMS 297  
302 LGKKYVGTGNGADPTISGLGAWTKPTQWNSNFFENFYGMELTKSPAGAVQWPKRD 361  
298 LGKRSVGTGNGADPTISGLGAWTKPTQWNSNFFENFYGMELTKSPAGAVQWPKRD 357  
362 GAGAGTTPDAH-DPSKSHAFPMLTDLALRMDPDYKISRITYENPDFAAFKAMVKL 420  
358 GAGAGTTPDPGGRGS--PTPKATDLSRYDPIYETIRRWLEHPEELAEFAKAMVKL 415  
421 THDMGKPVRYLGPVQEDLWDDPIPDVSHPIVDNDIDGLKAKILESGITVSELVST 480  
416 IHDMGKPVRYLGPVQEDLWDDPIPDVSHPIVDNDIDGLKAKILESGITVSELVST 475

QY 481 ANASSTFNRSKRGANGARIRLAPQKDWVNNPQ-QLARVLRKLTGIEQDFNQASDN 539  
DB 476 ANASSTFNRSKRGANGARIRLAPQKDWVNNPQ-QLARVLRKLTGIEQDFNQASDN 535  
QY 540 KAVSLAHLIVLAGCGAVERAAKDAGHEVOYFNPGRADATAROTVAPALPAADGFR 599  
DB 536 IKVSADLVVGGCAIEKAAKAAGHNTVPTPTORTDASOBTQVDSFVLEPKADGFR 595  
QY 600 NYIKPEKYSAEMLVQRAQLLSUSAPKALVGMGVVIGTNTDSQGRGVFTNKPQOLSN 659  
DB 596 NYLKGKNPAPAEYMLDKANLLTSAPENVVGGVNLVGLNGLNTRKPLPVFTFEASSTN 655  
QY 660 DEFVNLIDNTKRRASDESDKVFGRDFTGVKMSGTVDLIFGNSNSELRAALAEVYCA 719  
DB 656 DEFVNLIDMTWEPSPADDTGYQGD-GSKVKVMTGSRVDLVFGNSNSELRALVEYVAD 714  
QY 720 DSEKVFQDYKAKAKVGLDRFDLK 745  
DB 715 DAQPFQDYKAKAKVGLDRFDLK 740  
RESULT 6  
AAW33809  
ID AAW33809 standard; Protein: 753 AA.  
XX AAW33809;  
DT 18-NOV-1998 (first entry)  
DE Alcaligenes (Deleya) aquamarinus catalase-64CA2.  
KW Catalase: epoxidation; hydroxylation; biosensor; paper bleaching;  
XX pasteurisation.  
XX Alcaligenes aquamarinus.  
XX W09800526-AI.  
XX 08-JAN-1998.  
XX 03-JUL-1997; 97NO-US15513.  
XX 03-JUL-1996; 96US-0674887.  
XX (RECO-) RECOMBINANT BIOTECALYSIS INC.  
XX Adhikary RS, Robertson DE, Senyal I;  
XX WPI; 1998-085953/08.  
XX N-PSDB; ANV0554.  
XX New bacterial catalases, related nucleic acid vectors and  
XX transformed cells - used as oxidising agents and for detecting or  
XX destroying hydrogen peroxide, e.g. in biosensors  
XX Claim 1; Fig 1; 35pp; English.  
XX The present sequence is of the Alcaligenes aquamarinus catalase-64CA2.  
XX Catalase-64CA2 may be used to catalyse oxidation reactions such as  
XX epoxidation or hydroxylation. The enzyme can also be used to detect or  
XX destroy hydrogen peroxide, e.g. in connection with glyoxylic acid  
XX production, biosensors, contact lens cleaning, pulp/paper bleaching and  
XX pasteurisation of dairy products. Antibodies raised against  
XX Catalase-64CA2 can be used to screen libraries for detection and  
XX identification of the gene. The present sequence is of the gene  
XX encoding Catalase-64CA2 can be used to identify related sequences.

XX Sequence 753 AA;  
Query Match 57.2%; Score 2290.5; DB 19; Length 753;  
Best Local Similarity 59.1%; Pred. No. 3.8e-193;  
Matches 433; Conservative 92; Mismatches 183; Indels 25; Gaps 8;









QY 525 LEG-----IQDPNQAOSDKAVSL-----ADLVLAGC-----AQVEKAAKNDAGHDVQV 569  
 Db 511 LAGHPARTAEKSHRLDLGGLLNKQKQATPALASCHFSLAARHSHKSPASKALPC 570  
 QY 570 PFNPGRADATQDVEAFEALEPAADGPRNYIKPEHKVSAEMLVDRAQLLSAPENT 629  
 Db 571 WNSQMASAT-----IKSKSTRPRKSCSTKPS-----SADRPRND 608  
 QY 570 ALGCMARVLTGTYGQUGHTNFKPQGLSNDFVNLDLNWKWASDSKVFEGDRFT 689  
 Db 609 GLGWRPARVGNPHLPGHVFDRIGVLNDFVNLDDNNKYNVPTDGG--IYEIRDRFT 666  
 QY 690 GEVKSNGSTRVDLTGFSNELRALAEVYGCADSEKFKYDVKAKAVKMDLRDPL 744  
 Db 667 GEVMTATRVLDLTFGTAQDNOEFAADQDNOEKFVRDFTINAWVKVWNRDPL 721

RESULT 12  
 ABG25057  
 ID ABG25057 standard; Protein; 533 AA.  
 AC ABG25057;  
 XX  
 XX 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #25048.  
 XX  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 XX Homo sapiens.  
 XS WO200175067-A2.  
 PN 11-OCT-2001.  
 XX  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2001; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 XX (HYSB-) HYSBQ INC.  
 PA Dmanac RT, Liu C, Tang YT;  
 PI WPI: 2001-639362/73.  
 DR N-PSDB; AAS9244.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 XX Claim 20; SEQ ID No 55416; 103pp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC genes. As antibodies against (I), (II) are useful in detecting or  
 CC quantitating a polypeptide in tissues. (I) is useful in medical  
 CC as a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to assess the type and frequency of products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 XX Sequence 533 AA;  
 Query Match 31.6%; Score 1266; DB 22; Length 533;  
 Best Local Similarity 60.9%; Pred. No. 9.4e-103;  
 Matches 255; Conservative 55; Mismatches 95; Indels 14; Gaps 6;  
 QY 328 TPTONSNEPHEULGYSVELTSPAGAYQKFKQAGAGCTIDAHDPKSHAPKMLUTDL 387  
 Db 126 TPTONSNEPHEULGYSVELTSPAGAYQKFKQAGAGCTIDAHDPKSHAPKMLUTDL 183  
 QY 388 ALRMDPDYKISRISYENPDEFADAFKAKMTKLTHRDGPKYRLVGPVQEDLWQDPI 447  
 Db 184 TLAFDFPEKISRFRFLNQDQNFPAFAWFKLTHRDGPKYRIGVPFKREDLWQDPL 243  
 QY 448 PD-VSHPLVDENDTEGLKAKILESGUTVSELYSTAWASSTPRMSKGGANGARTLAP 506  
 Db 244 POPIYNP--TEODIIDLKFAIDSGLSVSELYSVAMASITFRGDKGGANGARLALMP 301  
 QY 507 Q-KDMEVNPQOLARVLTLEGIQDPNQAOSDKAVSLADLVLAGCAGVEKAAKNDAGH 565  
 Db 302 QTRIGDVTNRAPI-RALPVEKIQESGK-----SLADLVLAGCAGVEKAAKNDAGH 353  
 QY 566 EVQVPPNPGRADATQDVEAFEALEPAADGPRNYIKPEHKVSAEMLVDRAQLLSA 625  
 Db 354 SIHVPPFAGPRVDAQDQTAIGMFELLEPIADGFRNTRARLDVSTTESLLDKAQLUTLA 413  
 QY 626 PENTALVGNHVRVLTGTYGQUGHTNFKPQGLSNDFVNLDLNWKWASDSKVFEGR 685  
 Db 414 PENTALVGNHVRVLTGTYGQUGHTNFKPQGLSNDFVNLDLNWKWASDSKVFEGR 473  
 QY 686 DFTKTEVKNSTGYDLITGSENELRALAEVYGCADSEKFKYDVKAKAVKMDLRDPL 744  
 Db 474 DRETGEVKFTASRADLVFGSNLYRAVRYASSDAHEKFKVDFAVAVKVMNLRDPL 532

RESULT 13  
 ABG08823  
 ID ABG08823 standard; Protein; 273 AA.  
 AC ABG08823;  
 XX  
 XX 13-FEB-2002 (first entry)  
 DT  
 XX Novel human diagnostic protein #8814.  
 XX  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 XX  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US08631.  
 PF 31-MAR-2001; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 XX (HYSB-) HYSBQ INC.  
 PI Dmanac RT, Liu C, Tang YT;  
 XX  
 XX WPI: 2001-639362/73.  
 DR N-PSDB; AAS73010.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT amino acid sequences. ABG00010-ABG30377 represent novel human



XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN W0200175067-A2.  
 XX 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Dmanac RT, Liu C, Tang YT;  
 XX WPI: 2001-639362/73.  
 XX DR N-PSDB; AAS73015.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX PS Claim 20: SEQ ID NO 39187; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC mapping. (II) is useful as probes for the detection of genetic disorders  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC diagnosis, forensics, gene mapping, identification of mutations  
 CC disorders involving abnormal gene expression and its applications in  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of DNA and products dependent on DNA and  
 CC amino acid sequences. AAG00010-AAG0377 represent novel human  
 CC Nucleic acid sequences of the invention.  
 CC No The sequences described herein do not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 201 AA:  
 Query Match 1148; Score 455; DB 22; Length 201;  
 E-Value Similarity 4.33e-137; Percent Identity 100;  
 Matches 105; Conservative 26; Mismatches 42; Indels 74; Gaps 4;  
 QY 468 LESCLTVSELVSTAMASATFRNSDKGGANGARIRLAPQKDEVNNPQQLARVLKLEG 527  
 DB 4 LPSLFTFFRLSS-----SNDILRGDKGGANGARIALPQ-----TRTG 43  
 QY 528 IQEDFNQASDNKAVSLADLLVLACGAGYKAKDAGHEVQVPFNPGRADATAE----- 582  
 DB 44 -----DYTRQPIRVC 54  
 QY 583 TDVEAFEALPAACGFRNYIKPEHKVSAEMLVDRAQLLSLAPETALVCGHVRVLTGY 642  
 DB 55 TVTEMPELLPEIADGFRNRYARLVDSYTESLLIDKAOQLTAPETALVCGHRYLGANF 114  
 QY 643 DCSGHQVTHKPCQGLSNDFPNVLLDLYTKWASDSKRVPEGROFTGEVWSGTRVDLI 702  
 DB 115 DQSKNGVPTDRGVLSNDFFNLLDMRYEMKATDSNELFEGROBEVGVKFTFASRADLV 174

Qy 703 FCSMSEL 709  
 Db 175 FCSMSVL 181  
 Search completed: October 7, 2003, 19:42:23  
 Job time : 63.1769 secs



Result No.	Score	Query Match Length	ID	Description
1	4002	100.0	745	US-08-674-8878-8
2	4002	100.0	745	US-08-951-844-8
3	4002	100.0	745	US-09-412-347-8
4	2544	63.6	740	US-08-418-782-7
5	2544	63.6	740	US-08-418-782-21
6	2544	63.6	740	US-08-228-662-7
7	2544	63.6	740	US-08-652-219-7
8	2544	63.6	740	US-08-852-219-21
9	2421.5	60.5	726	US-08-313-185-49
10	2421.5	60.5	726	US-08-439-185-11
11	2421.5	60.5	726	US-08-439-185-16
12	2397	59.9	729	US-08-313-185-51
13	2397	59.9	729	US-08-459-499-14
14	2397	59.9	729	US-08-459-499-14
15	2331	58.2	726	US-09-082-614A-50
16	2290.5	57.2	753	US-09-338-352-8115
17	2290.5	57.2	753	US-08-674-8878-6
18	2290.5	57.2	753	US-08-951-844-6
19	2278.5	56.9	735	US-09-412-347-6
20	2278.5	56.9	735	US-08-313-185-48
21	2278.5	56.9	735	US-08-439-499-9
22	2270.5	56.7	735	US-08-439-499-9
23	1954	48.8	652	US-08-562-547-46
24	1954	48.8	652	US-08-459-499-17
25	1954	48.8	652	US-09-082-614A-53
26	1760.5	44.0	731	US-08-439-499-17
27	1760.5	44.0	731	US-08-459-499-15



Qy 721 SEEKFKYDFVKANAKYMDLDRFDLK 745  
Db 721 SEEKFKYDFVKANAKYMDLDRFDLK 745

## RESULT 3

US-09-412-347-8  
; Sequence 8, Application US/09412347  
; Patent No. 5658733  
; GENERAL INFORMATION:  
; APPLICANT: Robertson, Dan E.  
; APPLICANT: Sanyal, Indrajit  
; APPLICANT: Adhikari, Robert S.  
; TITLE OF INVENTION: CATALASES  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; COUNTRIES: India, India  
; STREET: 4225 Executive Square, Suite 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MOLECULE TYPE: protein  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/412,347  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/674,887  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Ph.D., Lisa A.  
; REGISTRATION NUMBER: 38,347  
; REFERENCE/DOCKET NUMBER: 09015/002001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/678-5070  
; TELEFAX: 619/678-5070  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 745 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: Internal  
US-09-412-347-8  
Query Match 100.0%; Score 4002; DB 4; Length 745;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 745; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MENHKSQSYTYWTKGCKPTGGSQKAGSGGTNRDWWPMLNLTGLRQHSLSDPN 60  
Db 1 MENHKSQSYTYWTKGCKPTGGSQKAGSGGTNRDWWPMLNLTGLRQHSLSDPN 60  
Qy 61 DPDFDAEYFKKLDLAAYKKDLAALMTDSQDWPDYDGHYGPFTIRMAHSAGTYRIGDG 120  
Db 61 DPDFDAEYFKKLDLAAYKKDLAALMTDSQDWPDYDGHYGPFTIRMAHSAGTYRIGDG 120  
Qy 121 RGGGSGSORFAPLNSPNDANLDKARLLWLPKQYKRTKSWADLITGNVALTWGPF 180  
Db 121 RGGGSGSORFAPLNSPNDANLDKARLLWLPKQYKRTKSWADLITGNVALTWGPF 180  
Qy 181 KTFGAGGADWPEEDYVYGAATWGLDKYEGCDRELENPLGAVQWGLTYNPEGNG 240  
Db 181 KTFGAGGADWPEEDYVYGAATWGLDKYEGCDRELENPLGAVQWGLTYNPEGNG 240  
Qy 241 KPDPFAAARDIRETEFGAMNDDEETVALIAGGHTFGKTIAGAAAEKYVGREPAAAGIEEM 300  
Db 241 KPDPFAAARDIRETEFGAMNDDEETVALIAGGHTFGKTIAGAAAEKYVGREPAAAGIEEM 300

Qy 301 SLGKNNTYCTGCAOTITISGLEGAWTKPTQWSNNFFENLFGYEWELTKSPAGAYOWPKP 360  
Db 301 SLGKNNTYCTGCAOTITISGLEGAWTKPTQWSNNFFENLFGYEWELTKSPAGAYOWPKP 360  
Qy 361 DGAGAGTIPDAHPDPSKSHAPFMTTDLALMDPDYKTSRRTYENPDEADAFKAKYKL 420  
Db 361 DGAGAGTIPDAHPDPSKSHAPFMTTDLALMDPDYKTSRRTYENPDEADAFKAKYKL 420  
Qy 421 THRMGPKRYLGPVEPQEDLLMODIPDVSHPLVDENDIEGLKAKLIESGLTVSELVST 480  
Db 421 THRMGPKRYLGPVEPQEDLLMODIPDVSHPLVDENDIEGLKAKLIESGLTVSELVST 480  
Qy 481 ANASASTFNSDKRGKGANGARIRLAPQKDWENNPQQLARVLTLEIGIDEPNQAQSDNK 540  
Db 481 ANASASTFNSDKRGKGANGARIRLAPQKDWENNPQQLARVLTLEIGIDEPNQAQSDNK 540  
Qy 541 AVSLADILVLACGACVYKAKAGAGHEVQVFPNPGRADATASQTDVEAFALPAAGFRN 600  
Db 541 AVSLADILVLACGACVYKAKAGAGHEVQVFPNPGRADATASQTDVEAFALPAAGFRN 600  
Qy 601 YIKPEHKVSAEMLVDRAQLLSLAPETALVGGHRYLTGYDSQHGVTNKPQGLSND 660  
Db 601 YIKPEHKVSAEMLVDRAQLLSLAPETALVGGHRYLTGYDSQHGVTNKPQGLSND 660  
Qy 661 FVNLDLMTWNRSDSKVPCGROPYKGVKWSGTWDLIFGSNSLRLALRVGACD 720  
Db 661 FVNLDLMTWNRSDSKVPCGROPYKGVKWSGTWDLIFGSNSLRLALRVGACD 720  
Qy 721 SEEKFKYDFVKANAKYMDLDRFDLK 745  
Db 721 SEEKFKYDFVKANAKYMDLDRFDLK 745

## RESULT 4

US-08-418-782-7  
; Sequence 7, Application US/08418782  
; Patent No. 5658733  
; GENERAL INFORMATION:  
; APPLICANT: Cockerill, Franklin R.  
; APPLICANT: Kline, Bruce C.  
; APPLICANT: DNL James R.  
; TITLE OF INVENTION: Detection of Isoniazid Resistant Strains  
; TITLE OF INVENTION: of M. Tuberculosis  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwegman, Lundberg & Woessner  
; STREET: 3500 IDS Center  
; CITY: Minneapolis  
; COUNTRY: MN,USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT PATENT DATA:  
; APPLICATION NUMBER: US/08/418,782  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Woessner, Warren D.  
; REGISTRATION NUMBER: 30,440  
; REFERENCE/DOCKET NUMBER: 150.141051  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-339-3351  
; TELEFAX: 612-339-3306  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 740 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear



```

QY 600 NTYKPHYVAKEMVDRAQLLSLGAPEMTALVGSNRVLTGTYWGSOHGVYTNKPGQSLN 659
DB 596 NYLKGKNEPLAEVYMLDLKANLLTSLAEPNTVLVGLGRVLYGANYKRLPLGYTFTEASESITN 655
QY 660 DFFVNLDLNTKWRASDSKVFEGRDKFTGEVWSGTRVDLITGNSLSLALAEVYGA 719
DB 656 DFFVNLDMGITWESPADDTGTQSKD-GSGKVMYTWGSRVDLITGNSLSLAEVYVGD 714
QY 720 DSEKIKVDFVYAKAVYDLDRDLK 745
DB 715 DAQKPKYDFVYAAVMKYNLDREDFVR 740

RESULT 6
US-08-228-662-7
; Sequence 7, Application US/08228662
; Patent No. 5925275
; GENERAL INFORMATION:
; APPLICANT: COCKERILL, FRANKLIN R.
; APPLICANT: KLINE, BRUCE C.
; TITLE OF INVENTION: DETECTION OF ISOMIAZID RESISTANT STRAINS
; TITLE OF INVENTION: DETECTION OF M. TUBERCULOSIS
; NUMBER OF SEQUENCES: 7
; NUMBER OF SEQUENCES: 7
; ADDRESS: SCHWEGAN, LUNDBERG & WOESSNER, P.A.
; CITY: 3500 IDS CENTER
; STATE: MINNEAPOLIS
; COUNTRY: USA
; ZIP: 55402
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/228,662
; PRIORITY: 18-434-1594
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WOESSNER, WARREN D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 150.123US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; INFORMATION FOR SEQUENCE 7:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
US-08-228-662-7

Query Watch 63.68; Score 2544; DB 1; Length 740;
Best Local Similarity 64.38; Pred. No. 9.5e-222;
Matches 480; Conservative 88; Mismatches 168; Indels 10; Gaps 8;

QY 2 ENHKGSGSYNTNGKPTTGSLKQSGAGGCTKMRDWPNNGLILRSHSLSDPN 61
DB 3 EDDPTTITTTGASNG-CFVGHMKYVPEGGG--NQDWPRLKMLVLRNFAVADPMG 59
QY 62 PDYDPAEKFKDLDAVKKDLAALMTSDWNPADYGHYGPFTIRMAHNSAGTYRDCGR 121
DB 60 AAFDPAAEVATIDVATLTDEEVWITSQFWNPADYGHYGLFIRMAHNAAGTYRIDGR 119
QY 122 GGGGSGSQRAPLNSWPNAMLOKARLLWPIKQYGRKISWADMLITGNVALETGFK 181
DB 120 GAGGQGRAPLNSWPNAMLSKARLLWPPVKYTKKLSWADLITVFNAGCALSGMFK 179
QY 182 TFGFAGGADWPEEDYVNGAETKLGDKRDEDELENPLGAVQNGLITVYVPGNKG 241

```

```

DB 180 TFGFGGRVQWEPDE-VYMGKATWLGDRYSCKDLNPLAAYVQMLITVYVPGN 238
QY 242 PDPTAAARDTRETGCRWANDDETVALIAGCHTGTGTCDAADAKYVYHSDAAGIGENS 301
DB 239 PDPMAAAADTRETFRAMANDVETAAALVGGHTFGTHTGAPAD-LVGPPEAAAPQMG 297
QY 302 LGWNTYTGTHGADTITSLGSGAMTKPTPONSNEFFENLGYEMELTKSPAGAYQKPKD 361
DB 298 LGWASGTGTGTDKDAYTGYVWNTPTKWNDSFLEILYGYEMELTKSPAGAYQKPKD 357
QY 362 GAGAGTIPDAH-DPSKSHAPPLMTDLALMDPDYKLSRBYENPDOPADAFANAYKL 420
DB 358 GAGAGTIPDPFGGGRS--PTMLATDLSLAVDPIYERITRWLHPLELADEFANAYKL 415
QY 421 THRDMPGKRYVLPGEVPEQEDLNQDPIPOVSHPLVDENDIEGLAKLTESGLTVSELVST 480
DB 416 THRDMPGKRYVLPGEVPEQEDLNQDPIPOVSHPLVDENDIEGLAKLTESGLTVSELVST 475
QY 481 AWASASTSRNSDKRGANGARILAPQKDWYNNPO-OLARYLKTTLGGLOEDFNOASDN 539
DB 476 AWASASTSRNSDKRGANGARILAPQKDWYNNPO-OLARYLKTTLGGLOEDFNOASDN 535
QY 540 KAVSLADLITVLAGCAVEKAAKADAGEVYVPPENFGGRADATADTDFEALPADGFR 599
DB 536 IATSEADVLVLAGCAVEKAAKADAGENITVPTPTGTDASQEDTDFEALPADGFR 595
QY 600 NTYKPHYVAKEMVDRAQLLSLGAPEMTALVGSNRVLTGTYWGSOHGVYTNKPGQSLN 659
DB 596 NYLKGKNEPLAEVYMLDLKANLLTSLAEPNTVLVGLGRVLYGANYKRLPLGYTFTEASESITN 655
QY 660 DFFVNLDLNTKWRASDSKVFEGRDKFTGEVWSGTRVDLITGNSLSLALAEVYGA 719
DB 656 DFFVNLDMGITWESPADDTGTQSKD-GSGKVMYTWGSRVDLITGNSLSLAEVYVGD 714
QY 720 DSEKIKVDFVYAKAVYDLDRDLK 745
DB 715 DAQKPKYDFVYAAVMKYNLDREDFVR 740

RESULT 7
US-08-852-219-7
; Sequence 7, Application US/08852219
; Patent No. 5925275
; GENERAL INFORMATION:
; APPLICANT: COCKERILL, Franklin R.
; APPLICANT: Kline, Bruce C.
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Selection of Isomiazid Resistant Strains
; TITLE OF INVENTION: Of M. Tuberculosis
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Mueeting, Raasch & Gebhardt, P.A.
; STREET: 119 No. 5922575th Fourth Street, Ste. 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,219
; FILING DATE: 07-May-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sandberg, Victoria A.
; REGISTRATION NUMBER: 41,287
; REFERENCE/DOCKET NUMBER: 230.00010130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1226

```



Db 476 AWAASPSRSGKRGANGKICPOVQWENPDGDLAKVITLREZIQSFNKAAPN 535  
 QY 540 KAYSALADLYLAGACAGVEKAKDAGHEVOVFPNPGRADATASOTDVEAFEALEPAADGR 599  
 Db 536 IKSVPADLVYLAGCAAEKAAKAGHNIIVPPTGRDASOBQDVSFSAVLEPKADGPR 595  
 QY 600 NYIKPHKVAZEMLVDRQAQLISAPEMTALVCGNRVLTVYDQSGHGVFTNKPCOLSN 659  
 Db 596 NYLKGPNPLAEKALNDRALNLTGAEVLYVJGGLRYGRTKPLGYFTENBSSELIN 655  
 QY 660 DFFVNLDMGVTWPPSPADDTYQKQ-GSKRVMKTVGSRVLDVFGSNSELRALVETGAD 714  
 Db 656 DFFVNLDMGVTWPPSPADDTYQKQ-GSKRVMKTVGSRVLDVFGSNSELRALVETGAD 714  
 QY 720 DSEKTVKADVPVANKVXMDLRDPLK 745  
 Db 715 DAQPTVQDTPVAANDKVNLRDVR 740

## RESULT 9

US-08-313-185-49  
 ; Sequence 49, Application US/09311185  
 ; Patent No. 5871912  
 ; GENES/FEATURE INFORMATION:  
 ; APPLICANT: Heym, Beate  
 ; APPLICANT: Cole, Stewart  
 ; APPLICANT: Young, Douglas  
 ; APPLICANT: Zhang, Ying  
 ; APPLICANT: Honore, Nadine  
 ; APPLICANT: Telenti, Amalio  
 ; APPLICANT: Bommer, Thomas  
 ; APPLICANT: Kohn, Michael  
 ; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance  
 ; TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis  
 ; NUMBER OF SEQUENCES: 66  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ; ADDRESSEE: Dunner  
 ; STREET: 1300 I Street, N.W.  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; SOFTWARE SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE RELEASE: #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA  
 ; APPLICATION NUMBER: US/08/313,185  
 ; FILING DATE: 12-OCT-1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meyers, Kenneth J.  
 ; REGISTRATION NUMBER: 25,146  
 ; TELEPHONE: (202) 408-4000  
 ; TELEPHONE: (202) 408-4000  
 ; TELEFAX: (202) 408-4400  
 ; INFORMATION FOR SEQ ID NO: 49:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 726 amino acids  
 ; TYPE: amino acid  
 ; STRANDNESS: single  
 ; MOLECULE TYPE: Peptide  
 ; US-08-313-185-49  
 ; Query Match 60.5%; Score 2421.5; Db 2; Length 726;  
 ; Best Local Similarity 63.6%; Pred. No. 1.2e-210;  
 ; Matches 467; Conservative 78; Mismatches 172; Indels 17; Gaps 7;  
 ; 12 YWTVWGCKPPTGSSIKQSGAGGTNRDWFNKLNLGILQLQHSISDPNDPDDFAEKP 71

Db 8 HNTATGKCFQGGHDSAGAGTTRDAMPNQLRVLLIQAHSNRNPLGEDFYDKRES 67  
 QY 72 KLDAAVKKDLAAWTDSDQMRANDYGYGVFTFRANHSNAGTYRTGDRGGSGGQRF 131  
 Db 68 KLDYGLAKDLKALTESOPWPDWAGSTAGLFRAMRGAGTYRSIDRGAGRGQRP 127  
 QY 128 APINSPONANDKARLLIWPITKQYKIKSWADLMITCNVALETMGFTFGFAGRAD 191  
 Db 132 APINSPONANDKARLLIWPITKQYKIKSWADLMITCNVALETMGFTFGFAGRAD 191  
 QY 128 APINSPONANDKARLLIWPITKQYKIKSWADLMITCNVALETMGFTFGFAGRAD 191  
 Db 132 APINSPONANDKARLLIWPITKQYKIKSWADLMITCNVALETMGFTFGFAGRAD 191  
 QY 192 VNEPDEYVNGATKEMGLSKYSGRELENPGLGAVOMGLTYVNPPEGPKPDPIAARDI 251  
 Db 188 VNEPDLVNGDEKAWLTHRPEA--LAKAPLGATMGLIYVNPPEGDISCEPLSAAAI 245  
 QY 252 RETFGMWNDEETVALIAGGTFGKTGHGAADAEKTVGRPAAAGTEEMSIGKMTGTG 311  
 Db 246 RATFGMWNDEETVALIAGGTFGKTGHGAADAEKTVGRPAAAGTEEMSIGKMTGTG 311  
 QY 312 HGADTITISGLEAGWTKTPQMSNFFENLFGVMELTSPAGAYOMKPKDGCAGCTPDA 371  
 Db 305 VGADATISGLEAVVMTQTPQMSNYFFENLFGVMEVQTPSPAGAIQEPAYD--APETIDP 362  
 QY 372 HDPEKSHAPMLTDLALMDPPDYBKISHRYTENPDEPAOFAAKAKLITHRMGPVRY 431  
 Db 363 FQPSKEKFTMLVDTLAFDFEFKISRFLNDPQANFARAKMLTHRMGPVRY 422  
 QY 432 LGPEVQEDLIWQDIPD-VSHPLVDENDIEGLKAKILSGLYSELVSTWASASTFRN 490  
 Db 423 IGPEVPKEDLIWQDLPQIYNP--TQDIDLFIADISGLSYSELVSVANASASTFRG 480  
 QY 491 SKRPGCANGARILAPKDWENVPNQLARVLTLEGIQDFNQASDNKAVSLADLIIV 550  
 Db 481 GURKGGANGARIALMPORDVW--AARVALPVLKIQESKRA-----SLADIIVL 531  
 QY 551 AGCAGVEKAKDAGHEVOVFPNPGRADATASOTDVEAFEALEPAADGPRNTKPHKVA 610  
 Db 532 AGVYGVKSAAGLSIRYFPAFGVDAQDQTDIEMFELLPADQDPRNTKPHKVA 591  
 QY 611 ERMVDRAQLISAPEMTALVCGNRVLTVYDQSGHGVFTNKPCOLSNDFVNLDM 670  
 Db 592 ERMVDRAQLISAPEMTALVCGNRVLTVYDQSGHGVFTNKPCOLSNDFVNLDM 670  
 QY 671 KWRASDESKVPEGRDKTKGVKWSGTVDLLIFGNSSELRALAEVYGCADSEKPVDFV 730  
 Db 652 ENKATSEKLPFGEDRGRETGEVFTASRADLVFGNSVLRVAEYTAASDAEHEKVDYF 711  
 QY 731 KANKVXMDLRDPL 744  
 Db 712 AARVXMDLRDPL 725

## RESULT 10

US-08-459-499-13  
 ; Sequence 13, Application US/08459499  
 ; Patent No. 5871912  
 ; GENES/FEATURE INFORMATION:  
 ; APPLICANT: Beate  
 ; APPLICANT: Cole, Stewart T.  
 ; APPLICANT: Young, Douglas B.  
 ; APPLICANT: Zhang, Ying  
 ; TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods  
 ; TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis Resistant to Isoni  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ; ADDRESSEE: Dunner  
 ; STREET: 1300 I Street, N.W.  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315







Db 590 DVSTFSLIDKAAQLTLTAEPTVYVGNRVLTWFGNSQNGVFTDKRGVLTDFPML 649  
 QY 666 LQNTWFAWSDSKVFCEDRGTGKVSCTRVLTGNSLRLALARYTGCDSERKF 725  
 Db 650 LDMRYEMKPTDANELFECGRDLTCEVKYATRADLVTGNSVLRALARYTGCDSERKF 709  
 QY 726 VYDFVAAWAKYMDLDRDLK 745  
 Db 710 VYDFVAAWAKYMDLDRDLQ 729  
 RESULT 13  
 US-08-459-499-14  
 ; Sequence 14, Application US/08459499  
 ; Patent No. 5871512  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reym, Beate  
 ; APPLICANT: Chou, David T.  
 ; APPLICANT: Young, Douglas B.  
 ; APPLICANT: Zhang, Ying  
 ; TITLE OF INVENTION: Nucleic Acid Probes, Sequences, and Methods  
 ; TITLE OF INVENTION: for Detecting Mycobacterium Tuberculosis Resistant to Isoniazid  
 ; TITLE OF INVENTION: Amended  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ; ADDRESS: 1300 I Street, N.W.  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.3  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE: 08/08/459,499  
 ; FILING DATE NUMBER: 1995  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/875,940  
 ; FILING DATE: 30-APR-1992  
 ; PRIOR APPLICATION DATA:  
 ; FILING DATE: 07/929,206  
 ; FILING DATE NUMBER: 1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/029,555  
 ; FILING DATE: 11-MAR-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meyers, Kenneth J.  
 ; REGISTRATION NUMBER: 25,146  
 ; TELECOMMUNICATIONS INFORMATION:  
 ; TELEPHONE: 202-408-4000  
 ; TELEFAX: 202-408-4400  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; LENGTH: 729 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Peptide  
 US-08-459-499-14  
 Query Match 59.9%; Score 2397; DB 2: Length 729;  
 Best Local Similarity 62.8%; Pred. No. 2e-208;  
 Matches 465; Conservative 82; Mismatches 171; Indels 22; Gaps 9;  
 QY 11 TFWNTGCKPTFGSLKQSGAGTGRKQWPKMLNLTILRHSSLSLDPNDPRDYAEF 70  
 Db 7 THNTLSTGCKPFGHGIDSGAGTASBWWPNQLRVLLNQNSNPLGDFDYAEF 66

71 KELD-LAAYKKDLAALMTSDWMPADYGHYGPFFIRMAHISAGTTRIGDRGGSGSSO 129  
 Db 67 SKLDTLSALAGDLSQDTPWPAWGLTFIRMAHAGTGYTSDIGRGAGRG 126  
 QY 130 REAPLANSQNDLTKARLLWPTKQYGRKISWADMLTGNVALETGKPTGCPAGR 189  
 Db 127 REAPLANSQDVTSLDKARLLWPTKQYGRKISWADLILAGNVALENGSPTGFCAGR 186  
 QY 190 ADWPEEDYVCAETFWLGDRELENPGLGAVQMGSLIYVNGPCKPDPFAAR 249  
 Db 187 EDWPEFDLVNMGDEKAWLTHRPEA--LAKAPLGATMDLIYVTPGPNHSGEPLSAA 244  
 QY 250 DIBETGRMANDEETVALIAGHTGKTHGAADNKYGRPRACIGTEMSLGHKTYG 309  
 Db 245 ATRATGTGNGMDEETVALIAGHTLGTGTPA-AASHVGADPEAPIEAGLGNASSTG 303  
 QY 310 TGHGAOTITSGLEGATKTPQMSNFFENLPGYEMELTSPAGAYOMPKDGAAGTIP 369  
 Db 304 SGVANDITSGLEVYVQTFQMSYFFENLFTZWGTQSPAGUQFENY--APDITP 361  
 QY 370 DAHDSKSHIA--PMLVTDLALMDPYEKTSRRYENPDEPADAKAYKLTHTMGCP 427  
 Db 362 DFPDQKRXKXKPTMLVTDLALRFDPEKTSRRFLPDQAFNEAFARAFKLTHTMGCP 421  
 QY 428 KYRILGPVPOEDLMDQDPDVSPLVD--ENDIEGLKALILESGTLVSELVSTAWASA 485  
 Db 422 KARYTGPVEKDELMDQLP--QPLVQPTQEDILNLAATAASGLISERSVYAKSR 478  
 QY 486 STFNENSKRGKGANGARILAPQKDMVNNIQLARVLATLEGIEDFENAGSDNKAVSLA 545  
 Db 479 STFRGGDKRGKGANGARLALAPQDMVN--AVAAVLVPLEEIQKTINR-----SLA 529  
 QY 530 DIIYLAGVVGIEQAAAAAVSIHVPPPPGVDARHHCOTDIEFSLLEPACDTRNRYARL 589  
 QY 546 DELVACACGYEKAAGDAGHEVQVVPNGRADATAGTQDVEAFPALEPADGFRWYIKRP 605  
 Db 606 HKVSAREMLVDRQALSLSPENTALYVGNRVLTNYDSQHGVTNKPQLSNDFFVNL 665  
 QY 590 DVSTFSLIDKAAQLTLTAEPTVYVGNRVLTWFGNSQNGVFTDKRGVLTDFPML 649  
 QY 666 LQNTWFAWSDSKVFCEDRGTGKVSCTRVLTGNSLRLALARYTGCDSERKF 725  
 Db 650 LDMRYEMKPTDANELFECGRDLTCEVKYATRADLVTGNSVLRALARYTGCDSERKF 709  
 QY 726 VYDFVAAWAKYMDLDRDLK 745  
 Db 710 VYDFVAAWAKYMDLDRDLQ 729  
 RESULT 14  
 US-09-082-614A-50  
 ; Sequence 50, Application US/09082614A  
 ; Patent No. 6124098  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Reym, Beate  
 ; APPLICANT: Chou, David T.  
 ; APPLICANT: Young, Douglas  
 ; APPLICANT: Zhang, Ying  
 ; APPLICANT: Monore, Nadine  
 ; APPLICANT: Telenti, Amalio  
 ; APPLICANT: Bodmer, Thomas  
 ; TITLE OF INVENTION: Rapid Detection of Antibiotic Resistance  
 ; TITLE OF INVENTION: in Mycobacterium Tuberculosis  
 ; NUMBER OF SEQUENCES: 60  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ; ADDRESS: 1300 I Street, N.W.  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315  
 ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/082,614A  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIORITY DATA:  
 ; APPLICATION NUMBER: US 08/313,185  
 ; FILING DATE: 12-OCT-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meyers, Kenneth J.  
 ; REGISTRATION NUMBER: 25,146  
 ; REFERENCE/POCKET NUMBER: 02356-0068-00000  
 ; COMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 408-4400  
 ; TELEFAX: (202) 408-4400  
 ; INFORMATION FOR SEQ ID NO: 50:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 729 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Peptide  
 ; US-09-082-614A-50

Query Match 59.9%; Score 2397; DB 3; Length 729;  
 Best Local Similarity 62.8%; Pred. No. 2e-208;  
 Matches 465; Conservative 82; Mismatches 171; Indels 22; Gaps 9;  
 QY 11 TYNATGCKGPTGGSLKOSAGCKYKRWKNNMLNGLILROHSLSDNDPDYAFEP 70  
 Db 7 THTLSLTKGCPFHGGHDSAGTAGSDWNLQVLLNQHNSNPLCGEDFYRKEF 66  
 QY 71 KKLD-LAAVKDILAALMTDSQWMPADYGHYGPFFIRMAHISAGTYRIGDGRGGGSGSQ 129  
 Db 67 SKLDYSAIGDGLKALDTPQWMPADWSYVGLFIRMAHISAGTYRISIDGRGGAGQQ 126  
 QY 130 REPANLSPDNOKARLLIWPYKQYGRKISWADMLTGNVALETMGKTFPSAGCR 189  
 Db 127 REPANLSPDYNLSDKARLLIWPYKQYGRKISWADLFILAGNVALENSGFRTFEGSAGR 186  
 QY 190 ADVPEEDYVCAETENIGDKRYEGDRELENPILGAVQKGLIYVNPSPGNCKPDPITAAAR 249  
 Db 187 EDWPEEDYVNGWDEKANTIRPEPA--LAKAPIGATEMOLIYVTPGNSHSEPLSAAA 244  
 QY 250 DIERETGKANDREETVALTAGSTFKTHGCAADKRYKQVREPAAGIETMSLQHWKNG 309  
 Db 245 ATRATFGNKNDEETVALTAGSTLTKTHGPA--AASHVGADEPAAPTEAGLGNWASSTG 303  
 QY 310 TGHGAQDITISGLGEMATKPTQNSNFFENLFGYEWELTKSPAGAYQKPKDGAAGATIP 369  
 Db 307 GAHAQDITISGLGEMATKPTQNSNFFENLFGYEWQVTKSPAGATQFPAVD--APDITP 361  
 QY 370 DAHDPSKSHA--PQMLTDLAKMDQDVEKISRYETNGDFADAKAKYKLTUSDMG 427  
 Db 362 DPEDPSKRXKPTMLTDLARDFEPEKISRRLDQAFNEAFARAKYKLTURDMGP 421  
 QY 428 KYRTIAGPEVQEDLWDPTIDPVSPLVD--ENIDELGKALIESGLTYSLSVSTANASA 485  
 Db 422 KARYTIGPEKEDLWDQDPLP---QPLVQPTQEDINLKAATASGLTYSLSVSTANASA 478  
 QY 486 STFNESDEKANGARILALAKQKENVNPOOLARVLTKLEGTQEDFNQASDNKAVSLA 545  
 Db 479 STFRGDKRGANGARILALAPQMDVN--AAVARVLPVLEELQKTNKA-----SLA 529  
 QY 546 DLITVAGCAGVEKAAAGHEVQVVPENPGRADATAGQDVFAPALPAGDFRNYIKPE 605  
 Db 530 DLITVAGVWLEQAAAAARVSIHVPPPGVDAHQDQDTEMPSLLEPTADGFRNAR 589  
 QY 606 HYSASMLVDKRALSLSEKQWMTALGVGNVLTGQVSGQVETNKKPQOLSNDFPNI 665

Db 590 DVNSTTTSLLIDKAQULLFLAPEMTVVGGHVRVLGTWFGNSQNVFTKFCVLTSTOFFANL 649  
 QY 666 LDLATNTPWASDSQDVPGRDPTGVKNGSGTVDLTGNSKSLPALAATYVGCASDERE 725  
 Db 650 LDMRYEMPTDDANLEFEGRDRTGTGVKTYATRADLVTFGNSVLSALAEVYACSDAHEK 709  
 QY 726 VDEPVAWAKYMDLDLDRFLK 745  
 Db 710 VKDFVAMVKYMWLQREDFLQ 729  
 RESULT 15  
 US-09-328-352-8115  
 ; Sequence 8115, Application US/09328352  
 ; GENERAL INFORMATION:  
 ; PATENT NO. 6562958  
 ; APPLICANT: GARY L. BRESTON et al.  
 ; TITLE OF INVENTION: BACULOVIRUS AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 8115  
 ; LENGTH: 26  
 ; MOLECULE TYPE: Peptide  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-8115

Query Match 58.2%; Score 2331; DB 4; Length 726;  
 Best Local Similarity 61.8%; Pred. No. 1.9e-207;  
 Matches 450; Conservative 83; Mismatches 177; Indels 18; Gaps 7;  
 QY 19 KCPFTGSGSLKQ--SAGGCTKRWKNNMLNGLILROHSLSDNDPDYAFEFKDLDA 76  
 Db 14 KCPFTSCHNSKPTVYGGGTANLHNWENLQVLLNQHNSNPLDKDNTQEFKLDYY 73  
 QY 77 AYVKDIAALMTDSQWMPADYGHYGPFFIRMAHISAGTYRIGDGRGGGSGSQRFAPLNS 136  
 Db 74 ALKADIKNVLDTSQWMPADMTGCTIRLHAGNYTRNGDGGGAGQQRFAPLNS 133  
 QY 137 WPDNANLDKARLLIWPYKQYGRKISWADMLTGNVALETMGKTFPGAGGRADWPE 196  
 Db 134 WPDNANLDKARLLIWPYKQYGRKISWADLFILAGNVALENSGFRTFEGSAGREDWPE 193  
 QY 197 EDYTWGAETENIGDKRYEGDRELENPILGAVQKGLIYVNPSPGNCKPDTAAARDITETG 256  
 Db 194 NOVWMDKEMLAHRNSEA--LAGSMLAATFEMGLIYVNPSPGASGPRKSAAPTITATFG 251  
 QY 257 RNANNDREETVALTAGSTFKTHGCAADKRYKQVREPAAGIETMSLQHWKNTYTGHGADT 316  
 Db 252 NNAMDEEETVALTAGSTLTKTHGAAPAD--HVQADPEGATTEQGFQWANSYGTGVGKDA 310  
 QY 317 ITSGLEGAMATKPTQNSNFFENLFGYEWELTKSPAGAYQKPKDGAAGATIPDANDPSK 376  
 Db 311 ITSGLEVWSTQVTPQNSNFFENLFGYEWQVRSFAGALQWAAQ--AAALIPDPDPDSI 368  
 QY 377 SHAPFMTDLAKMDQDVEKISRYETNGDFADAKAWYKLTURDMGPKRYLGPVE 436  
 Db 369 KSKPTMLTDLTDRDFEPEKISRRLDQAFANAFARAKYKLTURDMGPKRYLGPVE 428  
 QY 437 PQEDLWDQDPTIDPVSPLVDENDIBGLKALIESGLTYSLSVSTANASASTFRNSDKGG 496  
 Db 429 PABDLWDQDLPAPAS--ATPSSNANADAKAYVALGSLAGELSVLSANASASTFRGDKRG 487  
 QY 497 ANGARILALAKQKENVNPOOLARVLTKLEGTQEDFNQASDNKAVSLADLVLAGAGV 556  
 Db 488 ANGARILALSPQRENVNKK-----KAVETLTKIEELKASTQ-----LSIADLVLAGNKG 537  
 QY 557 EKAAGAGHEVQVVPENPGRADATAGQDVFAPALPAGDFRNYIKPEHKYKSEDLVD 616  
 Db 538 EQAQAAGFNITVPPAGRVGNLQSGQVTSFOLLGLADGFRNKKQGVNPAEVLVD 597

QY	617	RAQLLSAPENTALVGNRVIGTNYDQSHGVFNKPGQLSNDFFVNLDLNTRKASD	676
Db	598	RAQULUFRPBLTALUGLRLVLTNWDGSGRGFTQQVGLSTDFIALLDMSNVPYD	657
QY	677	ESDKVEGRDFKTGEVKNSGTRVDLIFGNSLRLAELVYGCADSEKFKYDFYKAMKV	736
Db	658	STSEVEGCKDRKSGTVKFTATRNLDLVFGNSILRLAELVYQAQDGKEKFKYQDFVAATKV	717
QY	737	WLDLRFDL	744
Db	718	WLDLRFDL	725

Search completed: October 7, 2003, 19:47:37  
Job time : 24.8825 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: October 7, 2003, 19:45:48 Search time 48.241 Seconds  
441000 Alignments/CS  
2449.330 Million cell updates/sec

Title: US-09-884-889-8

Perfect score: 4002

Sequence: 1 MENHKGSGSYNTNGKCT.....VKDFVRAKMYNDLOREDFLK 745

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 587654 seqs, 156212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published\_Applications\_AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US05\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4002	100.0	745	10	US-09-884-889-8
2	2290.5	57.2	753	10	US-09-884-889-8
3	273.5	6.8	291	10	US-09-870-501-1
4	255	6.4	250	14	US-10-027-559-6
5	223.5	6.4	176	10	US-09-712-017-8
6	166	4.0	176	10	US-09-738-626-6164
7	121	3.0	867	10	US-09-738-626-6164
8	121	3.0	976	15	US-10-128-714-8481
9	117	2.9	673	15	US-10-156-761-8864
10	113	2.8	331	12	US-10-259-165-16
11	113	2.8	331	12	US-10-259-165-360
12	112.5	2.8	1104	15	US-10-156-761-9790
13	111	2.8	728	15	US-10-156-761-9352
14	106	2.6	1104	15	US-10-156-761-9352
15	110	2.7	1208	15	US-10-156-761-13251

Sequence 67, Appl	15	110	2.7	1477	12	US-10-193-764-67
Sequence 47, Appl	16	110	2.7	2516	9	US-09-811-511-2
Sequence 47, Appl	17	110	2.7	2516	15	US-10-242-056-47
Sequence 3135, Ap	18	110	2.7	2152	15	US-10-128-714-3135
Sequence 3135, Ap	19	109.5	2.7	2209	15	US-10-128-714-3135
Sequence 10145, A	20	109.5	2.7	1247	9	US-09-815-243-10145
Sequence 5962, Ap	21	109	2.7	904	10	US-09-738-626-5962
Sequence 12011, A	22	108.5	2.7	493	15	US-10-156-761-12011
Sequence 9789, Ap	23	107.5	2.7	1281	12	US-10-156-761-1281
Sequence 12011, A	24	107.5	2.7	1281	12	US-10-156-761-1281
Sequence 9789, Ap	25	106	2.6	431	15	US-10-043-173-4
Sequence 4, Appl	26	106	2.6	431	15	US-10-150-762-4
Sequence 4, Appl	27	106	2.6	431	15	US-10-150-762-4
Sequence 7036, Ap	28	106	2.6	1097	12	US-10-032-585-7036
Sequence 7965, Ap	29	105.5	2.6	7746	15	US-10-156-761-7746
Sequence 14403, A	30	105	2.6	720	15	US-10-156-761-14403
Sequence 3481, Ap	31	105	2.6	918	15	US-10-128-714-3481
Sequence 9361, App	32	105	2.6	426	15	US-10-043-173-4
Sequence 9361, App	33	105	2.6	426	15	US-10-043-173-4
Sequence 11, Appl	34	104	2.6	742	14	US-10-077-111-11
Sequence 11, Appl	35	102.5	2.5	811	11	US-09-864-636A-586
Sequence 586, App	36	102	2.5	811	11	US-09-864-636A-588
Sequence 586, App	37	101.5	2.5	611	11	US-09-952-267-11
Sequence 11, Appl	38	101.5	2.5	622	15	US-10-156-761-12898
Sequence 12898, A	39	101.5	2.5	1248	10	US-09-738-626-4814
Sequence 4814, Ap	40	101	2.5	3239	15	US-10-032-585-7036
Sequence 2, Appl	41	101	2.5	3239	15	US-10-032-585-7036
Sequence 2, Appl	42	100.5	2.5	366	10	US-09-885-339-4
Sequence 4, Appl	43	100.5	2.5	855	15	US-10-156-761-11464
Sequence 11464, A	44	100.5	2.5	900	15	US-10-156-761-12858
Sequence 12858, A	45	100.5	2.5	3256	10	US-09-919-172-98

#### ALIGNMENTS

RESULT 1  
US-09-884-889-8  
; Sequence 8, Application US/09884889  
; Patent No. US20020102680A1  
; GENERAL INFORMATION: CORPORATION  
; APPLICANT: ROBERTSON Dan  
; APPLICANT: SANYAL, Indrajit  
; APPLICANT: ADHIKARI, Robert  
; TITLE OF INVENTION: CATALASES  
; FILE REFERENCE: DIVER100-4  
; CURRENT APPLICATION NUMBER: US/09/884, 889  
; CURRENT FILING DATE: 2001-06-19  
; PRIORITY DATE: 2001-06-19  
; PRIOR FILING DATE: 1999-10-05 US/09/412, 347  
; PRIOR APPLICATION NUMBER: US 08/951,844  
; PRIOR FILING DATE: 1997-10-16  
; PRIOR APPLICATION NUMBER: US 08/674,887  
; PRIOR FILING DATE: 1996-07-03  
; NUMBER OF SEQ ID NOS: 8  
; SEQ ID NOS: 1-8  
; SEQ ID NO: 6 PatentIn version 3.0  
; LENGTH: 745  
; TYPE: PRT  
; ORGANISM: Microscilla furvescens  
US-09-884-889-8

Query Match 100.0%; Score 4002; DB 10; Length 745;  
Matches 745; Conservative 0; Mismatches 0; Gaps 0;  
QY 1 MENHKGSGSYNTNGKCTPTGSGKQSGAGGCTKNRWPNMNLGLILRHSSLDPN 60  
DB 1 MENHKGSGSYNTNGKCTPTGSGKQSGAGGCTKNRWPNMNLGLILRHSSLDPN 60  
QY 61 DPFDYAEFFKLDLAAYKDKLAALMTSDQMPADYGHYGPFFIRMAHNSAGTYRGD 120  
DB 61 DPFDYAEFFKLDLAAYKDKLAALMTSDQMPADYGHYGPFFIRMAHNSAGTYRGD 120



```

; SEQ ID NO 1
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Hordeum vulgare L. Haruna-ni.jyo
US-09-870-501-1

Query Match
Best Local Similarity 6.8%; Score 273.5; DB 10; Length 291;
Matches 94; Conservative 99; Mismatches 95; Indels 115; Gaps 10;

Oy 102 PFETMAHSAAGTYRIGCGSGSOREFAPLNSPNDKARLLWPIKQYGRKI 161
Db 32 PHLRLAHADGATYDV-NRTFGANGSRVREYETHGSNAGLKATDLLEPIAKH-PKI 89
Oy 162 SHADMLILGNVALETMGKTFGAGGADWVEPEEDYMGAEFTWIGDKRYEGDRELM 221
Db 90 TADLHQLAGVAVYEGGFVEPIGRND----- 119
Oy 222 PLGAVQWGLIYNPSPGNGK-PDPTAARDIRETFGRMANDEETVALLAGHTGKTHG 280
Db 130 -----SSVCPR--EGRLPAKKGAPHLRIDFYRMGLTKDKDIVAL--SGCHSLGKAHP 167
Oy 281 ADAXEYVGRPAAGIEHMSLGMKNYVYVGHGADITTSGLRGAWTPTQWSNPPFNI 340
Db 168 E-----RSGDGMATNDPLFNDSTF--- 188
Oy 341 FGYEWELTSPAGAYOMKPDGAGATTIPDAHDPSKSHAPMLTTDLARMDPYEKISR 400
Db 189 ----LEILK-----GESGLLKLPTDKALDDPEFRRYVE 219
Oy 401 RYVENDEPDAFAKAYKILHRDMGPKRYVIGPEVQEDL 441
Db 220 LYANDEDYFKDAVSAHKLKLS-ELGFTPHSSGASTKSDV 258

RESULT 4
US-10-027-559-6
; Sequence 6, Application US/10027559
; Publication No. US20020144307A1
; ORGANISM: Zea mays
; APPLICANT: Simion; Carl
; TITLE OF INVENTION: Maize Defense-Inducible Genes and Their
; FILE OF INVENTION: Use
; CURRENT APPLICATION NUMBER: US/10/027,559
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 60/243,120
; PRIORITY DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Zea mays
US-10-027-559-6

Query Match
Best Local Similarity 6.4%; Score 256; DB 14; Length 250;
Matches 88; Conservative 99; Mismatches 95; Indels 126; Gaps 12;

Oy 65 DYAEFEKKIDIAAVKKDLAALMTDSQWNPADYGHYGPFFTRMAHSAGTYRIGDGRG 124
Db 11 EYSEVDR-----ARRKRLALAKS-----CAPMLRLAHWSAGTFDV--SRPTG 55
Oy 125 GSGSORFAPLNSPNDKARLLWPIKQYGRKISWADLMILTGNVALETMGKTFG 184
Db 56 PFTGTHQSELAHGANAGLDITAVALLEPIKEEF-PILSYADFYVLGAVVAVETGPEIP 114
Oy 185 FAGGADWVEPEEDYMGAEFTWIGDKRYEGDRELENPLGAVQMGLIYNPSPGNGKPD 244
Db 115 FHPQREKPKPP-----PEG--RLPDA 134
Oy 245 IAAARDIRETFGR--MANDEETVALLAGHTGKTHGAEADAKKYVGREPAAAGIEHSLG 303

; SEQ ID NO 1
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Hordeum vulgare L. Haruna-ni.jyo
US-09-870-501-1

Query Match
Best Local Similarity 6.8%; Score 273.5; DB 10; Length 291;
Matches 94; Conservative 99; Mismatches 95; Indels 115; Gaps 10;

Oy 102 PFETMAHSAAGTYRIGCGSGSOREFAPLNSPNDKARLLWPIKQYGRKI 161
Db 32 PHLRLAHADGATYDV-NRTFGANGSRVREYETHGSNAGLKATDLLEPIAKH-PKI 89
Oy 162 SHADMLILGNVALETMGKTFGAGGADWVEPEEDYMGAEFTWIGDKRYEGDRELM 221
Db 90 TADLHQLAGVAVYEGGFVEPIGRND----- 119
Oy 222 PLGAVQWGLIYNPSPGNGK-PDPTAARDIRETFGRMANDEETVALLAGHTGKTHG 280
Db 130 -----SSVCPR--EGRLPAKKGAPHLRIDFYRMGLTKDKDIVAL--SGCHSLGKAHP 167
Oy 281 ADAXEYVGRPAAGIEHMSLGMKNYVYVGHGADITTSGLRGAWTPTQWSNPPFNI 340
Db 168 E-----RSGDGMATNDPLFNDSTF--- 188
Oy 341 FGYEWELTSPAGAYOMKPDGAGATTIPDAHDPSKSHAPMLTTDLARMDPYEKISR 400
Db 189 ----LEILK-----GESGLLKLPTDKALDDPEFRRYVE 219
Oy 401 RYVENDEPDAFAKAYKILHRDMGPKRYVIGPEVQEDL 441
Db 220 LYANDEDYFKDAVSAHKLKLS-ELGFTPHSSGASTKSDV 258

RESULT 4
US-10-027-559-6
; Sequence 6, Application US/10027559
; Publication No. US20020144307A1
; ORGANISM: Zea mays
; APPLICANT: Simion; Carl
; TITLE OF INVENTION: Maize Defense-Inducible Genes and Their
; FILE OF INVENTION: Use
; CURRENT APPLICATION NUMBER: US/10/027,559
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 60/243,120
; PRIORITY DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Zea mays
US-10-027-559-6

Query Match
Best Local Similarity 6.4%; Score 256; DB 14; Length 250;
Matches 88; Conservative 99; Mismatches 95; Indels 126; Gaps 12;

Oy 65 DYAEFEKKIDIAAVKKDLAALMTDSQWNPADYGHYGPFFTRMAHSAGTYRIGDGRG 124
Db 11 EYSEVDR-----ARRKRLALAKS-----CAPMLRLAHWSAGTFDV--SRPTG 55
Oy 125 GSGSORFAPLNSPNDKARLLWPIKQYGRKISWADLMILTGNVALETMGKTFG 182
Db 52 GPGNGSRIRREHETHGANSGLKATDLCEGVYAKH-PKITYADLYQLAGVAVETGCPD 110
Oy 183 FGFAGGADWVEPEEDYMGAEFTWIGDKRYEGDRELENPLGAVQMGLIYNPSPGNGK- 241
Db 111 IVTFPGKID-----SNVCPGKRL 129
Oy 242 PDPTAARDIRETFGRMANDEETVALLAGHTGKTHGAEADAKKYVGREPAAAGIEHMS 301
Db 130 PDAKQGFQHLRDYFVRKMSDKDIVAL--SGHTLGRAHPE----- 168
Oy 302 LGMKNYVYVGHGADITTSGLRGAWTPTQWSNPPFNI-FGYEWELTSPAGAYOMKPD 361
Db 169 -----RSGDGMATNDPLFNDSTFVR-----LEILK----- 194
Oy 362 GAGACTIPDAHDPSKSHAPMLTTDLARMDPYEKISRYENPDEPDAFAKAYKIL 421
Db 195 -----GESGLLKLPTDKLLEDPEFRRLV-ELYADADEAFFRYAESHHKLKLS 240
Oy 422 HRDMCP 427
Db 241 ELGFNP 246

RESULT 6
US-09-734-017A-54
; Sequence 54, Application US/09734017A
; Patent No. US2002014242A1
; GENERAL INFORMATION:
; APPLICANT: Lorchl, Jens

```

```

; APPLICANT: Benz, Andreas
; APPLICANT: Reider, Thomas
; APPLICANT: Reider, Andreas
; APPLICANT: Cirpus, Peter
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; APPLICANT: Schmitt, Ralf-Michael
; TITLE OF INVENTION: The genus from Physcomitrella patens encoding proteins involved
; in the synthesis of amino acids, vitamins, cofactors, nucleotides and
; nucleosides
; TITLE OF INVENTION: synthesis of amino acids, vitamins, cofactors, nucleotides and
; nucleosides
; FILE REFERENCE: BASF-NAE-1331-99-US
; CURRENT APPLICATION NUMBER: US/09/734, 017A
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/171,100
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patentin Ver. 2.1/WordPerfect
; SEQ ID NO 54
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Physcomitrella patens
; US-09-734-017A-54

Query Match 4.0%; Score 161; DB 10; Length 176;
Best Local Similarity 30.3%; Pred. No. 1.7e-06;
Matches 54; Conservative 22; Mismatches 59; Indels 44; Gaps 7;

QY 102 PFFIRAHISACTYRIGDGRGGGSGRFPAPLNSWPDNANLDRKLLLPKIKYGRKI 161
DB 34 PFLRAHWSGSTD--QESYTGUGPLSTFQBELAGANGADJLVNLIQPIKQF--PEL 91
QY 162 SNADMLILGNVALETMGFTFGAGRADYVMEPEEDYMGAEIEMLDGRYEGDRELEN 221
DB 92 SYADFTYLAGVAVYEVGTGPTIPFHPGRKD-----HET-----CIVRG 129
QY 222 PIGAVONGLIYNEPGRMGKPDPIAARDIRETFGR--MANDEETVALIAGHTFGKTH 279
DB 130 RL-----PDATKGLDHLICVFTKQGLTDKDIYVL--SGATILGRCH 169

```

# RESULT 7

```

; Sequence 6164, Application US/09738626
; Sequence 6164, Application US/09738626
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: NAKAGAWA, HIROSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: OKUGI, HAKUHIKO
; APPLICANT: TOSHI, TOSHI
; APPLICANT: SENOH, AKIHITO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 6164
; LENGTH: 867
; TYPE: PRT

```

```

; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6164

```

```

Query Match 3.0%; Score 121; DB 10; Length 867;
Best Local Similarity 20.0%; Pred. No. 0.097;
Matches 152; Conservative 94; Mismatches 271; Indels 242; Gaps 42;

```

```

QY 39 DWMPNNMLGILRQHSLSDPDPDYAEFEKKDLDAVKKDLAA--LMTDSQ---DWM 93
DB 209 DVMK-----GTUTHA--ETPDQDTSLVPLALYCHSSLAUKDIALETETKQGPWY 261
QY 94 PADYGHGPR-----FIMAHSSATYRIGDGRGGGSGSQ--REFAPLNSPDNAN 142
DB 262 HRNEGVAYPGKYQDFVPEFNAGAMENAGAVTIDQEVFASKATRYIE----- 311
QY 143 LDKARLLNPIKQ--TGRKIS--WADMLITGNVALETMGFTFGAGRADYVMEPEED 198
DB 312 --REATIIJELAHMFGDLVTQMMDLWL-----NESFATMAISQAE--EFTYN 360
QY 199 VYR---GAETEWGDKRYEGDRELENPIGVOMGLIYVNEPGRMGKPDPIAARDIRET 254
DB 361 TAWVTFANVEKSM---AYQDQ-----LPSTHPVFSDC----- 390
QY 255 FGRMANDEETVALIAGHTFGKTHGAA---DAERYVGRPEAAAGTEE--MSLGNKTYG 309
DB 391 -----TDIETVDONTDFGTITAK--GASVLKQLOAYVGREEFACVYRHPANIANGHA-- 440
QY 310 TCGHAGDTITSGLEGANTTPTKSNPNFE---NLFQVE----- 444
DB 441 ---SEFDLLGAJESGSDLSDMANOMLKTGINTLGAKFTTNGKTYSEVTOTGAAPG 497
QY 345 -----WELTKSPAGYQWPKPKDAGAGTIPDAHDPKSHAPPLMTD----- 386
DB 498 AGELETRHIAVLKTLVDSLSNRYANRYELDCSGASTVE--EIVCLQADPVLNDDLLY 556
QY 387 LALRMDPDYKTSIRRY--YENPDEPA-----AFAKAWYKLTHRDMGPKYTLGPEVPOE 439
DB 557 ALDLDLDD-----SRNFVDNIKDFSDPMPRTLVMGAAMKTRAGQW--KAR----- 601
QY 440 DLHWQDPIPVSHPLVDENIDELGKAKILDESGLTVSELVSTAWASASTFRNSDKRGA-- 497
DB 602 -----DFALVARGAAATEIAYLERLQAQATSNLAKYADPAPNAEAT---GMDLLADPL 653
QY 498 NGARIRLAPOKDEWNNPQOLARVLKTLGSIQDFNQASDKNAKAVSLADILYLAGACGY 557
DB 654 RGAR--SAPDPSDTQIAFTQALAKA--PLNDAADYFRD-----ILAG--NWE 695
QY 558 KAKADGHEVYVPPNGRADTAEDTQVAFPE--ALEPAD-----GFRNYIKPEHKVSA 610
DB 696 -----GLTVDPDLRWWALTALLARGDIENVEDATAELSRSSASTLASLRAGAVVT 749
QY 611 EEMLVDRQAQLSLSAPEMTALVQGM-----RVLGTYNIDGSH-----CV 649
DB 750 EE-----YKAAAYKHVTAVDUSGLSNLELRHKIEGLTFTGSSLELLQAYNEYFILDY 802
QY 650 FTKKPOLSNDFVNNLLQATNKRASDE-----SDKVFEG 684
DB 803 NANFSGEMAQI---VGLFPSSNNVSEGLARTDFELDG 838

```

# RESULT 8

```

; US-10-128-714-8461
; Sequence 8461, Application US/10128714
; Sequence 8461, Application US/10128714
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroszhkin, Alexey M
; APPLICANT: Eroszhkin, Sebastian W
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use

```



```

: FILE REFERENCE: 10182-018-999
: CURRENT APPLICATION NUMBER: US/10/128,714
: CURRENT FILING DATE: 2002-04-23
: PRIOR APPLICATION NUMBER: US 60/285,697
: PRIOR FILING DATE: 2001-04-23
: PRIOR APPLICATION NUMBER: US 60/287,066
: PRIOR FILING DATE: 2001-04-22
: PRIOR APPLICATION NUMBER: US 60/295,890
: PRIOR FILING DATE: 2001-06-05
: PRIOR APPLICATION NUMBER: US 60/303,899
: PRIOR FILING DATE: 2001-07-09
: PRIOR APPLICATION NUMBER: US 60/316,362
: PRIOR FILING DATE: 2001-08-31
: NUMBER OF SEQ ID NOS: 8603
: SOFTWARE: GenBank version 3.1
: SEQ ID NO: 8481
: LENGTH: 976
: TYPE: PR1
: ORGANISM: Aspergillus fumigatus
US-10-128-714-8481

Query Match
Best Local Similarity 3.04; Score 121; DB 15; Length 976;
Beat Local Similarity 19.24; Pred. No. 0.12;
Matches 160; Conservative 99; Mismatches 255; Indels 318; Gaps 45;

QY 130 REFAPL-----SWPDNANLDRKLLWPIKQYKRSIWADMLWLGVALETMCF 180
Db 7 REVYTPQAINALSFSPSEAEI-----AGRGVPTLRAL----- 41
QY 181 KTFGAGGRADVMEPEEDYNGAETWAGTADKRYEGDRELENPLAGVOMGLIYV-NPE-G 237
Db 42 ---GRANDTEAWPKRGA-WFQEVWGRK---DRSIE-----GLAWLDPPEAG 85
QY 238 PNGKPPD-----IAAADIRTFGRMANDEETVALTAGCHTFCKTH 279
Db 86 PDGTFPKLRLFGISGYSTAVTEWNLQGRPMRISSNGYV-----EINCIAAQRWQATK 140
QY 280 GAADAE-----KYVGRPAAGTEHSELGWKNYV-----TG 311
Db 141 RKGDKGLPPAGSEYTGQH-LAGGAGSIVLSTADGDLKFLRLRSTKRAVLSTVF 199
QY 312 HGADTITSGLEGAMTKPTPQNSNFENLGYEWELTKSPAGAYO---WKPK----- 360
Db 200 QNRMTIVAGYADSSIRFLDLSQLLRT-----SLRGAGGPKELLVMSVKLPDGI 254
QY 361 ---PGAGCTYFDANDS---KSHAPNPTDIALMDPD-----YEKISRYEMPOE 408
Db 255 VSGESAGEIRFDANKYLSIORLOGLADLDVAVSANGCTVYVSGAORTVYIKRGE 314
QY 409 FADAFKAKYKLTIR-----DMGPKV---RYLG----- 433
Db 315 KGEKKR-WAEMVRRYTHDVTAVYETHDISIVYSGGPDSPVPLPLREFGEHRK 373
QY 434 PPYPO-----EOLWQDIPDWSH-PLVDEMTGSKAKTL----- 468
Db 374 LSSLQVLPOLASSPSESLVNSFWDEYSIW---VSRGPTSPHDSLEGRILVGRVL 428
QY 469 ---ESGLTSEL-----VSTAWASATFRNSDKRG-----GANGAR 501
Db 429 IQGENITISAMLSADGKLIVATVSATKYVFSYRRKSDERGTLRTKLDVPAAFSKDGAR 488
QY 502 -IRLAPQDM-----EVNNFQOLARVL-----RTLEGI 528
Db 489 AVSISPDSEWLCVFPDSNATYLRVNPASSFOQJLPOLVQLKLDRAHTRHREKASHG 548
QY 529 QEDFNGA-----QSDNRKAVSLADLIVLAGC-AGVERAAKADAGEVOPVNPGRADTA 580
Db 549 LGDTERTVRCVYFSENSRIATGCG---LGGCIDTMLLDKDYKSSAVASNKRN-GAADSOD 604
QY 581 EOTDVAEALPANAOC-----FNTYKAPBKVSGEMCVD-----RAQLISAPET 629
Db 605 ESSD---DEDERVPIQDERQVYANDSPITPLASGVLLISFTRPSAPKMLITNODQSS 661

```

---

```

QY 630 --ALVGRMVLTNYGSGHGV--FTNKPQLSN-----DFVNLID--LNTKW 672
Db 662 QISLRGSELMALT--SHQLYVEDTLBKRJLSSNSRNKNKAYLPQFQKGVKDRACOLM 718
QY 673 RASDESDFVEGCRDFKTGEVAV-SGTRVDLLIFGNSLRLAAYVYGAUSDE 723
Db 719 DLSESRL-----WLYGTSLWLMFELNQDPPSSBEMKFAADGHE 758

RESULT 9
US-10-156-761-8864
: Sequence 8864, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: HOSOKAWA, JUN
: APPLICANT: HOSHIMIZU, HIROSHI
: APPLICANT: HOSHIMIZU, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: BATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: PRIOR APPLICATION NUMBER: US 2002-05-23
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 8864
: LENGTH: 673
: TYPE: PR1
: ORGANISM: Streptomyces avermitilis
US-10-156-761-8864

Query Match
Best Local Similarity 2.94; Score 117; DB 15; Length 673;
Beat Local Similarity 18.94; Pred. No. 0.15;
Matches 129; Conservative 75; Mismatches 247; Indels 230; Gaps 31;

QY 23 TGSGLASGAGGTGKRWHP-----NWLKGLILQHSSELDNDPDPDY 66
Db 5 TPKGLRLATGDTNPQNPETWQEDVRLMREANGTVNSVGIFSM--ALLESPGAYDF 62
QY 67 ABEFKKLDIA---AVKKDLAALMTDSODWV-----PADYGHGVFFIRMA-WHSA 112
Db 63 GWLDRLDLHLHNGRADITGTPVPAFMPYRAHPALPVTADGTRY-DFGSRGATCSN 121
QY 113 GYTELIDGRCGGSGGSGGSAEPLASNDPNDKRLALLWTLKQYKRTIS----- 162
Db 122 ADY-----RAAANITTLRA-----TRYAEHPALALMHVNEHVGVPVSACYCSCNAH 169
QY 163 ---WADMLITONVALETMGFKTFG--FAG-----GRA----- 190
Db 170 FRLMLDTDTYGTIDALNANGTAFWGRYAGDQINPRLPTVTPNGQOLDHKRFADATM 229
QY 191 -DYWVEDEYV-----KCAETMLGDKRY-----GDSSEL 219
Db 230 RENFVAERDLHRLAPQIPVTTFMTALSQCDSDYDWANGREVLDVINDHLLTIDGRTH 289
QY 220 ENPLGAVQMG-----LITVYEPGPNKDPDFAAADIRTFGRMANDEETVALI 269
Db 290 VNLAMADLRTSVASGAPMLLLEHSTGTVNMPHINAKP-----QMARNSLARVARG 343
QY 270 AGCHTFCK-THGADARKYVGRPAAGKTEHSLGHKVTYGHGADPTTSSLEGAKTKT 328
Db 344 SEGANEFWQSRGAKGKTHSSMLPHSGTD--TRYVREYVELGSSVSLAS-IRGTETRA 400
QY 329 PT-----QNSNFFENLFGYEWELTKSPAGYQMRPKDGAGCTIPDAPDPSKSHAPMLT 384
Db 401 DVAVLDMQSWAQLN-----QMRPSD-----HDPRE----- 428
QY 385 TDALINDPDTKISRRY-----YENPDEFADAF-----AKAYKLT-----HROMGP 427

```

```

DB 429 -----RADAFTEALDRILVDFEADSDASIPVYVFAULFTEAAGSNKLEIYENG 483
QY 428 K--VYLGFEVPEGLDWDPTDPVSHPLVDENDLEGLK-----KLESLGTVS 475
DB 484 TLVSYFSGVIDEIVHDGPPYAL-----RDVLGLAVFEESPLKDDIRVTFPGDS 537
QY 476 ELVSTANA-----SASTERSDKRGANGARIAPAKOWEYNNPQIARLVKLTLE 526
DB 538 ELGDGWSFVVRGAZETVYTAGTAPVATHRHFPGSANYIS-----TLTAQ 590
QY 527 QIGEDFNOAQDNKAVSLADL 547
DB 591 GLDALLGRAADARIAPRADL 611

RESULT 10
US-10-259-165-16
; Sequence 16, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Meng, Xun
; APPLICANT: Chang, Hui-song
; APPLICANT: R-1998, Steven P.
; APPLICANT: Glasebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiya
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Novartis, Nicholas
; APPLICANT: Ricovart, Nicholas
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 16
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-165-16

Query Match 2.8%; Score 113; DB 12: Length 331;
Best Local Similarity 20.4%; Pred. No. 0.11;
Matches 55; Conservative 26; Mismatches 97; Indels 92; Gaps 7;

QY 161 ISNDMLITGNVALETGKTFGFGAGRADVWEPEEDYVNGATEWLGDKRYEGDRELE 220
DB 123 VSCADVLAFARDALALVGNAYVQVGGRRD----- 153
QY 221 NPLGAVOMGLIYVNEPGNGKDPDIAARDIRETFGRMANDEETVALIAGHTFGKTHG 280
DB 154 -----GNVSVAGETNGMLPPSANYAOLNOMFAGKLTQAEKVAL--SGAHTIGVSHC 204
QY 281 ADAKRYVREPAAGACIEENSLKWKNYGTGCHGADTTTSGLEGAMTKPTQMSNNFENL 340
DB 205 SFSNRLYSSGPN-----GQDPSMDPSYVAALITQCPQ----- 239
QY 341 FGYEMELTSPAGYQMKPKDGAGCTIP--DAHDFSKSHAPF-----MLTTDLALR 390
DB 240 -----OCOP-----AACGVPMDAVTNPATFNATNYAAIVANGLVSSQALL 280
QY 391 MDPEYKISRNYENDEDFADAFKANKYL 420
DB 281 ADQTTAAQVGVYGTNNPDSFOTDFAAAMVKM 310

RESULT 12
US-10-156-9790
; Sequence 9790, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TAUTOSHI

```

```

DB 281 ADQTTAAQVGVYGTNNPDSFOTDFAAAMVKM 310

RESULT 11
US-10-259-165-360
; Sequence 360, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Meng, Xun
; APPLICANT: Chang, Hui-song
; APPLICANT: R-1998, Steven P.
; APPLICANT: Glasebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiya
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Novartis, Nicholas
; APPLICANT: Ricovart, Nicholas
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 360
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-165-360

Query Match 2.8%; Score 113; DB 12: Length 331;
Best Local Similarity 20.4%; Pred. No. 0.11;
Matches 55; Conservative 26; Mismatches 97; Indels 92; Gaps 7;

QY 161 ISNDMLITGNVALETGKTFGFGAGRADVWEPEEDYVNGATEWLGDKRYEGDRELE 220
DB 123 VSCADVLAFARDALALVGNAYVQVGGRRD----- 153
QY 221 NPLGAVOMGLIYVNEPGNGKDPDIAARDIRETFGRMANDEETVALIAGHTFGKTHG 280
DB 154 -----GNVSVAGETNGMLPPSANYAOLNOMFAGKLTQAEKVAL--SGAHTIGVSHC 204
QY 281 ADAKRYVREPAAGACIEENSLKWKNYGTGCHGADTTTSGLEGAMTKPTQMSNNFENL 340
DB 205 SFSNRLYSSGPN-----GQDPSMDPSYVAALITQCPQ----- 239
QY 341 FGYEMELTSPAGYQMKPKDGAGCTIP--DAHDFSKSHAPF-----MLTTDLALR 390
DB 240 -----OCOP-----AACGVPMDAVTNPATFNATNYAAIVANGLVSSQALL 280
QY 391 MDPEYKISRNYENDEDFADAFKANKYL 420
DB 281 ADQTTAAQVGVYGTNNPDSFOTDFAAAMVKM 310

RESULT 12
US-10-156-9790
; Sequence 9790, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TAUTOSHI

```

```

; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: HATTORI, MASAHIRA
; TITLE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PUBLICATION NO. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, HIKOSHI
; APPLICANT: HORIKAWA, HIKOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9790
; LENGTH: 1104
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9790

Query Match      2.8%; Score 112.5; DB 15; Length 1104;
Best Local Similarity 18.7%; Pred. No. 0.6;
Matches 136; Conservative 68; Mismatches 267; Gaps 31;

QY 99 HYGFPRFRMAHSAQTYR-----IGDGRGG-----GGSGSORFAPLNSW 137
DB 232 HFTAFDR-ARFSTGTMDRILNRPSTGSGTGFSGQRPVAGKAGAGYDEAPGSG- 289
QY 138 PUNANLAKARLLMLDKOKYKSKISWADMLTGNVALETWKFYF-----183
DB 290 -----PVNKVVG-1SYVSQAGAEALAVENPGRSFSEAVRDAAHAWR 328
QY 184 -----GFAGGRADWPEEDVYGAETWEL-----GDKRYEGRELENPLG- 228
DB 334 LGAVGVGGTAE-----DRTFTYALVHALHPNWSIDADRYGSDOKVHYVYGRHQAQY 389
QY 239 G-----LIVNPE-CPNKPDPPTAAARD-----IRETFGRMAVDREPT 265
DB 390 GFESCDVYRSVOLLTLOPRTGSDLAQSLAELARQNGIWRMLGASGTHWANGDPS 449
QY 266 VALIAGHGFYFKT-----HGAADA-----EKYV-----288
DB 450 PAKALGTRAFGCTDFDLGALUSLVTAATVPSQSLSAAGRPVLSVQORPSLDKYNRHY 509
QY 289 -----GREPAAGCTEEMSLGHKTYG-----309
DB 510 MFSVSNWAGAAETLEMSGADFALSOLARAAGEKETATFARSSQWNNFNIAAPSPS 569
QY 310 --TGRGADITYSG-----LEGATWK--TPTQSNFFENFEG-YEMELTSPAGAYQ-- 356
DB 570 ARVSGSGGLNSGTIANRKAQSGYGTGTAT-NGFVEGTGAQTWQVQENPAGUFAM 628
QY 357 -----WPKDKAGAGT-----TPADHP--SKSHAPMLTDLAL 389
DB 629 GGRDAALDRDAFHDSDGSWAFPGSGGDKSELNDEFSINPLYAVAGAPYKTQTVRA 688
QY 390 RMDPDYKISRRYENDEPDAFANAKMYKLTHRMGPKVRYLGPVEQDILWQDPI-P 448
DB 689 AMKLKSTPRGPIGNLDLR--KMSVFTYSLNGMTPOV-----PSRLVLSGLPFP 739
QY 449 DYSVELDNDISLKLALLESGLTYSSEL-----VSTAMASASTFNSDKRGANGARI 502
DB 740 RVEIRRHNDIEIRAQAKADAPYVRSLSKVDGRTSDRSNLPASEVED-----GGRIDTLL 795
QY 503 RLAPQDKQEVN-----NPOQL-----ARVILKTLRGIOE 530
DB 796 SATFNWAKADADAPSPRAGEQPTQIGVGPPTATVAPGGSTKIDTIALLSGGTGV 855
QY 531 DFNQAQSDNKAVSLADLLVLACGAYEKAADGAGHYOVQPNPGRADTAECTDVEAEA 590
DB 856 RFRVETPAGVATPAGTVSDGQAQITVTAADAAQ-----GYDVKVTVTSPTDSQVPA 912
QY 591 LEPAADG 597
DB 913 LTVAPG 919

```

```

RESULT 13
US-10-156-761-9552
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: HATTORI, MASAHIRA
; TITLE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PUBLICATION NO. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, HIKOSHI
; APPLICANT: HORIKAWA, HIKOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9552
; LENGTH: 728
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9552

Query Match      2.8%; Score 111; DB 15; Length 728;
Best Local Similarity 20.2%; Pred. No. 0.6;
Matches 166; Conservative 82; Mismatches 252; Indels 320; Gaps 48;

QY 5 KHGSGSYVNTGKCP--TGGSLKSGAGGCT-----KNRDWPMNLNIGILQHSSLS 57
DB 47 RHRRDLFFRT-AUGHRTFVQSGRTGVTVTGGTPTATOLGLNW-----LR 91
QY 58 DPNDFPDYAEFKLIDLAAYKDLAALMT-----DSQDMWPADYGHYGPFI 105
DB 92 NIADADLNWAG--ROLRIPLALPGLAGTVTTRAAVPHRFALNDTND-----GYTPYH- 142
QY 106 RWAKHSAGTYRIDGSGGSGSORFAPLNSHPDNALDKARLLMLPQIKYGRKTSWAD 165
DB 143 --DW--TV-----W-EREDVLAL-----BG-----YNE 161
QY 166 LMLTGLNVALETGKFTGFPAGGADVW--EPEDVTWGAE-----TEWLGDKRYE 214
DB 162 VLVQGDALHGVKFGFTGDEBLKARITPGAPHPMLNLNLSAFPDPVSOQLLDARAA 221
QY 225 GBELENPLGAKWGLIYVNPGE--PNKKDPTTAARDI-----RTTF 251
DB 212 LGRIANRLRELQMTGPTGVPFGFADRNAGACTPGGTWGFARPDMLDPRTEHF 285
QY 256 GRMA-----MNDHFVVALIAGHGFYFKT-----HGAADAQKVVGRPEAAAGIE- 299
DB 282 TRVAAAFYRIQDE-----MFGGASTRYKMDLLEHSGSGPVGVG--DAKKGVERALRAHP 335
QY 300 ----MSLCKN-----TYTGHGADTITSEL-----EGATKPTQSNFFEN 339
DB 336 GAWVYLLGHOHPRAIVDAVDKRMVLVGLCDRFKPYTDREADWHGTP-----Y 386
QY 340 LFGYEWEL-----TKSPAGAYQ-WKPKDGA---GAGTPDAHDPSKSHAPMLTDD 386
DB 387 APQSLNFWGFGHTTLGANTPDWASLYRDEWRTPTGSLAGVALLPEAD--NNPAAFALESE 444
QY 387 IALRMDPDYKISRRYENDEPDAFANAKMYKLTHRMGPKVRYLGPVEQDILWQDPI 446
DB 445 LAWE------CGLDLRWAFARWAS-----RYGGRD-PEAEAAW-- 477
QY 447 IPDVSHPLDNDITBGLKAKILLESGLTYSSELVSTAMASASTFNSDKRGANGARILAP 506
DB 478 -----DI-----LRTAYGTTRADNSGQADLPGFARPSLAA 509
QY 507 QR--DWEVN---NPOQ-----LAWVLKTLBGIO-----EDFNQAQSDNKAVSL-- 544

```

Db 510 TKAASPKRLRYRPEFEFPAELGELLKRVGRLGSSATRSATRDLLDVAQRLSNRSRVLLPQ 569  
 QY 545 -----ADLVLAGAGVEKA-----AKDAGHEVQVFPENGR--DATAQOTDVEA 587  
 Db 570 IRTAYEADTAERFORLTVGWLALMDLLEALLATDSHLLGRVWADARAGASAARORLA 629  
 QY 588 FEAL-----EPAD--GFNTYKPBKPVSAEMVDVRAQLLSLSPFMTALVGM--RV 637  
 Db 630 IDALSLLTVGTRAGADAGLADT-----ANENAGLVGLLFL 667  
 QY 638 LGTY-----DSQHGVEFTNKPQSLNDFYLLDLNTWK 672  
 Db 668 RMSTYFAELASASREGTRPK-----TDMFA-----LEDRW 698

## RESULT 14

US-10-156-761-11828

; Sequence 11828, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: IKEDA, SATOSHI

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 11828

; TYPE: PRT

; ORGANISM: Streptomyces avermitilis

US-10-156-761-11828

Query Match 2.84; Score 110.5; Db 15; Length 929;

Best Local Similarity 18.1%; Pred. No. 0.96;

Matches 137; Conservative 103; Mismatches 359; Indels 247; Gaps 36;

Matches 137; Conservative 103; Mismatches 359; Indels 247; Gaps 36;

QY 33 GGTNNEDWNNMLNLGILROHSSLD-----PNDPDDYAEFEKLDLAANKDLAL 85

Db 10 GGTGNNQPPFGAQCACPRNSNSHRKLPQEGGRAMARRGQSAPSRRRAAPRAAT 69

QY 86 MTSDDMWPAD---YGHYGPFFIRMAHSA--GYTRIGD--GRGGGSGSORFAPLNSW 137

Db 70 GSFSDYTGADDEERPG--GNADARASQSEFGRRRADAGRGSGGGGRRGP--GG 125

QY 138 FQAN-----LDKALLIAPTKYKGR-----ISKADMLITGNYA 174

Db 126 PGGPNQPGRCGRGVPKRRFIDYPRAGYGAARMLPSRWLVGICIAFTGSGNVAAGVA 185

QY 175 LETMGCTFTFGAGRADVREPEEDYTGAEEMLDGKRYEGDRELNPJAGVOMGLIY-- 232

Db 186 YALVGVFVYALTA-----KQNNYITWAGSQRVATGCTTRQIVY--ISQIPKAMRAY 238

QY 233 VNPEGNGKPPDIAARDIRETFGRMNDTEVALTAGCHTFKGTGAADAEKVG--- 289

Db 239 ISOENKTTETDSDVPKGIARAFNMA-----RCGETQG--GSTITQGVKNAM 285

QY 286 -----REPAANGTEENSLGKNTYGTGRGADYTSSL-----BGAW 325

Db 290 LDOQSOTISRKKEIFVAYGVKVPDAGTLAGTINSYVGRGATQAGARVFDIDNAK 345

QY 326 TKTPQSNHNFLENLEGEVEMELTSPAGAYONKPKDAGAGTIPDAHDPKSKHAPMLTT 395

Db 346 NUNPGCA-----FLANMLKATYYDP---AGATSLDPAATSEANRRKAKM 390  
 QY 386 DLALMDDDY---EKTSIRRY-----RNPDEFADAFKAKTKLTHRWGKRVYLGAPV 436  
 Db 391 QDTLDKEVEIHLDEATRNKYTELKPVENPNSARLSQGVLYV---DLAKAYLVNSEL 447  
 QY 437 POEDL-----IW-----QDPIQVSHPLVDENIEGLKAKILESGLITVSELS 479  
 Db 448 SADDLQARGQSYITTFDKKRYTELEDVADKAVRRKNIKPKQPDITDYVQPGASVDPPTG 507  
 QY 480 FAN-----ASATFNSDRGCGANAR--LAPOKQVYNPQOLANUKLTLEGTO 529  
 Db 508 AIKAIYGEDATKHTNNADQCAQVSGTFKPFVLAAMSGKKPOL----- 555  
 QY 530 EDFNQAQSDNAKVASLADLIVLAGCAGVEKAADAGHEVQVFPENGRADATABOTDAFE 589  
 Db 556 -DSSQSDERTVVS-----PKSLYSOKMLKLTEDYKGV---WTONGKE 596  
 QY 590 ALERPAQDAFNKYPEKHYSAEKI-----VDRAQQLSLSA----- 625  
 Db 597 WLO--TNDGGSYNAPSYKIDLREANRVNSANSAGFVQLGMDVGLDKVRDAALSAGILKTSLA 655  
 QY 626 -----PMTALVGMRYVG-----TNYDGS--QHG-- 649  
 Db 656 SANTPFSICTSDPAIRMACATATFASGKORQPYSEKRYTKDQVFIQUSIDIAKYDA 715  
 QY 650 FTNKPQSLSDNFYLLDLNTWKASDESQKVEGRDP--KTYGKWSKSTVDLLIFGNS 707  
 Db 716 FTKAVADNVDTVLKTVVQDGTGTAQ-----LTGREVAGTKGTT--DCKNSAMFVGTYP 767  
 QY 708 ELRALAEVGYCADSEKFPYKPYKAW 733  
 Db 768 QLSTSTWTRNDNSESKKROFLEMT 793

## RESULT 15

US-10-156-761-13251

; Sequence 13251, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OHURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; PRIOR FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; NUMBER OF SEQ ID NOS: 13109

; SEQ ID NO 13251

; TYPE: PRT

; ORGANISM: Streptomyces avermitilis

US-10-156-761-13251

Query Match 2.74; Score 110; Db 15; Length 1208;

Best Local Similarity 21.3%; Pred. No. 1.17;

Matches 94; Conservative 41; Mismatches 161; Indels 142; Gaps 17;

QY 6 HSGSSYTYNTTGGKCFP--TGGSLQASGGGKTRNRMWNLNLGILROHSSLSDDNDPD 63

Db 542 HGTATATYDGSKAPVQETSGQIVASAGAPV-----AHTA----- 578

QY 64 EGYAEFEKLDLAANKDLALMTDS-----QWHPADYGHYGPFFIRMAHSAQ 113

Db 579 FGLVKEASRYSLTVKVRQKAGANTAGLVQRLAAGTDPYPASVDSGTALRLA---PG 635

Query Match 2.74; Score 110; Db 15; Length 1208;  
 Best Local Similarity 21.3%; Pred. No. 1.17;  
 Matches 94; Conservative 41; Mismatches 161; Indels 142; Gaps 17;  
 QY 6 HSGSSYTYNTTGGKCFP--TGGSLQASGGGKTRNRMWNLNLGILROHSSLSDDNDPD 63  
 Db 542 HGTATATYDGSKAPVQETSGQIVASAGAPV-----AHTA----- 578  
 QY 64 EGYAEFEKLDLAANKDLALMTDS-----QWHPADYGHYGPFFIRMAHSAQ 113  
 Db 579 FGLVKEASRYSLTVKVRQKAGANTAGLVQRLAAGTDPYPASVDSGTALRLA---PG 635

```

QY 114 TYRIG---DGRGGGSGQREAPLNSWPDNANLKDRL---LMPKOKYGRKLSWADL 166
Db 636 TYSLSFDYVRGHHGADSLGLFLAAPTITLDRDRAITLDGSRLEVEADVDRKTETROL 695
QY 167 MLTGNVALETMGFTFGPAGGRADYWRPEEDVYWGAEPTWLGDKRYEGRELENPGLAV 226
Db 696 LR-----EYDRSGGGSOLF-----GAV 712
QY 227 OMGLIY-----VNEPGRKXPO-----PIAARDIRE-----TEGRMAMND 262
Db 713 QVPLTYDSVFAAPTPKVTEGSFEYRTVMRLGKPLLEVRGIREAVVQSGGTLTEGR-SRPP 771
QY 263 RETVALIAGGHFTGKTHGAADAEXYGRPEAPAGIEEMSLGNKNTYGTGCHGADTITSGLE 322
Db 772 LVDVGSFTGVYTVGRGAVLRLGIDGVEPTALQAQADAGAKALEFTDDSNGLT----- 827
QY 323 GAWTK-----PTQWGNFTYMLFCYENELTKSPAGYOKXKQG---AGACTIIDAH 372
Db 828 -AWNGTDONDRPLQIA-----TVNAADAARLRAGGRIDMTGTPTPTVY 872
QY 373 DPSKSHAPFM/TTDLALR 390
Db 873 DLSEGHGRIPDRDLTYR 890

```

Search completed: October 7, 2003, 19:59:27  
 Job time : 51.241 secs

**THIS PAGE BLANK (U)SP7C**

Result	No.	Score	Query	DB	ID	Description
1	2787.5	69.7	727	2	F95422	catalase (EC 1.11.1.6)
2	2759.5	89.0	735	2	B37663	catalase (EC 1.11.1.6)
3	2744.5	87.8	753	4	W50240	catalase (EC 1.11.1.6)
4	2714.5	57.8	753	4	W50240	catalase (EC 1.11.1.6)
5	2714.5	57.8	751	4	T43874	catalase (EC 1.11.1.6)
6	2714.5	67.8	751	4	T43878	catalase (EC 1.11.1.6)
7	2714.5	67.8	751	4	T43880	catalase (EC 1.11.1.6)
8	2714.5	67.8	751	4	T43881	catalase (EC 1.11.1.6)
9	2714.5	67.8	752	4	T43882	catalase (EC 1.11.1.6)
10	2714.5	67.8	753	4	T43882	catalase (EC 1.11.1.6)
11	2708.5	57.8	737	2	A07026	catalase/peroxidase (EC 1.11.1.7)
12	2691	67.2	737	2	A07026	catalase/peroxidase (EC 1.11.1.7)
13	2544	63.6	740	1	A40662	catalase (EC 1.11.1.6)
14	2541.5	63.5	746	1	A47685	catalase (EC 1.11.1.6)
15	2444.5	61.1	726	2	A10936	catalase (hydroperoxide EC 1.11.1.7)
16	2439.5	61.0	740	2	T45091	catalase (EC 1.11.1.6)
17	2424.5	60.6	728	2	G31237	hydroperoxidase HP (EC 1.11.1.7)
18	2423.5	60.5	725	1	C5E08E	catalase (EC 1.11.1.6)
19	2421.5	60.5	726	1	C5E08E	catalase (EC 1.11.1.6)
20	2417.5	60.4	726	2	C9E058	catalase (EC 1.11.1.6)
21	2417	60.4	724	2	B82186	catalase (EC 1.11.1.6)
22	2407	60.1	727	1	C5E08T	catalase (EC 1.11.1.6)
23	2401.5	60.0	720	2	T44562	catalase (EC 1.11.1.6)
24	2380.5	59.5	737	2	AC0403	catalase (EC 1.11.1.6)
25	2361.5	57.0	754	2	F75113	catalase (EC 1.11.1.6)
26	2356.5	56.6	741	2	A69239	catalase (EC 1.11.1.6)
27	2365	56.6	741	2	A69239	catalase (EC 1.11.1.6)
28	2233.5	55.8	720	2	S71130	catalase (EC 1.11.1.6)
29	2203.5	55.1	731	2	C98160	catalase (AB033631)

QY	193	WEPEDEYKAGATTEHL--GDKREGEORLENPFGAYOMGLIYVNPBGPKPDPIAAADT	251
DB	194		
DB	195		
DB	182	WHPEEDYVSEKEMLTGDKRYTSCORLENPFLAIVKGLIYVNEGPGDCKTDPKAAHDI	241
QY	252	RETFGRMAHDEETVALTAGHTFTGKHGAADAERYGPREPAAAGYEEMSLGHNKTYGTG	311
DB	253		
DB	254	RETFGRMAHDEETVALTAGHTFTGKHGAAGNDP-HVGPGRNPAETAGLGHOMTYGSG	300
DB	255		
DB	312	NGADITISLEGACANTPTQHSNKNVPGYFKELTSPDAGAYOKPKQCAAGCTTDA	371
DB	313		
DB	301	AGSDITISLEGACANTPTQWNGDFGFFLLFGYFMWLTSPAGAYOMQAVDPEKOLAPDA	360
QY	372	HDSEKSHAPMETDIALARNDPDYDEKISRRTYENDEPADAFAKAKYLTTHRMGPKVRY	431
DB	373		
DB	361	EDYKVPYVMTLTDIALARNDPDYDEKISRKHFNDEPADAFAKAMFKLLHRMDGPKARY	420
QY	432	LGPEVPEQDELWQDPIPDVSHPLVDNDETEGLAKLTESGLTSELVSTAWASASTFRNS	491
DB	433		
DB	421	LGPEVPEAFDEIWQDVPVDFYELTDA-EVEELKAKILDLSGLTSELVSTAWASASTFRNS	479
QY	492	DKRGKAGATIRLAPQKMEVNNPQOLARVLKLTGTEQDFQAOQSDNKVYASLDLIVLA	551
DB	493		
DB	480	DGRNGAKIRLAPQKMEVNNPQOLARVLENTQSOL-----DKRYASLDLIVLA	533
QY	552	SCAGYERAKADAGHYQVYVPPGRADATQATDVEAPALPAEDAFNGRTIKPEIKHYSAE	611
DB	553		
DB	534	GSAAVERAKAENGADVYVPPFPGQDQATQEQTDVEGFALVEPDSGFRTYOKKEYSVSPE	593
QY	612	EMLVQRAQLLISAPKWTALGVGNHVCYLTNYSQSGHYGKFNKPKQLSDNPFVLLDLNPK	671
DB	613		
DB	594	ELLIDIKQLLDLTAPEMTLALIGNRALGANYGSTQIGVFTDQVGTLLNDFEVLLDNGTE	653
QY	672	WRASDESDEKVFEGEDFKTEGVKMSGTVDLIFGSENLRAEAYGCGADSEKFKYDFVK	731
DB	673		
DB	654	KWPDVNI-LYFGKRGKRGGEVHTATRVOLVFGSNIILRAIATAYAQDNGKGFVEDFTA	711
QY	732	AWAKYMDLDRDFLX	745
DB	733		
DB	712	AWYKVMADRDFOLX	725
DB	713		
DB	714		

RESULT 3

JS0520

catalsae (EC 1.11.1.6) HPI - Bacillus stearothermophilus

N:Alternate names: catalase; peroxidase; hydroperoxidase I

N:Ontains: peroxides (EC 1.11.1.7)

C:Species: Bacillus stearothermophilus

C:Date: 30-Sep-1991 sequence\_revision 08-Nov-1996 #text\_change 03-Mar-2000

E:Accession: JS0520; A33955

E:Accession: JS0520; A33955

submitted to JGIPD July 1991

A:Reference number: JS0520

A:Accession: JS0520

A:Residues: 1-735 <TRA>

A:Molecule type: DNA

A:Residues: 1-735 <TRA>

J. Bacteriol. 171, 4871-4875, 1989  
A:Title: Cloning, nucleotide sequence, and expression in *Escherichia coli* of the BacI  
A:Reference number: A33955; MUID:89359121; PMID:2670897  
A:Accession: A33955  
A:Molecule type: DNA  
A:Residues: 1-731 <DOP>  
A:Sequence: The nucleotide sequence has been extensively revised in reference J80520  
A:Gene: C  
A:Gene: para  
C:Complex: homotetramer; two non-covalently associated iron protome IX groups per t  
C:Function: <CAT>  
A:Description: as catalase, catalyzes the dismutation of two molecules of hydrogen pe  
C:Function: <PER>  
A:Note: Active as peroxidase, uses hydrogen peroxide to oxidize donor compounds produ  
A:Note: Active with a broad spectrum of donor compounds, including larger ones which  
C:Superfamily: catalase HPI  
C:Keywords: chromoprotein; heme; homotetramer; iron; metalloprotein; oxidoreductase



F:101/Active site: His (distal axial ligand) \*status predicted  
 F:264/Binding site: heme iron (His) (proximal axial ligand) \*status predicted  
 F:315/Active site: Trp (tryptophyl radical intermediate) \*status predicted

Query Match 67.8%; Score 2714.5; Db 1; Length 735;  
 Best Local Similarity 68.7%; Pred. No. 5.1e-177;  
 Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

```

QY 1 MENKHSGSYTYWTFGGCKPFTGSGLSQAGGCTKRWMPNKLNLGLRHQSLSDPN 60
DB 1 MENO-----NRONAOCPEHGSVYTNQSS-NRTNKDMWPNQNLSTLHQRDKNPH 51
QY 61 DDPDYAEFFKLDLAAYKDLAALDSDSDWADYGYGTPFFIRMAHSGAGTYRGDC 120
DB 52 DEEFYAEFFKLDLAAYKDLAALDSDSDWADYGYGTPFFIRMAHSGAGTYRGDC 111
QY 121 RGGSGSGSORPAPLNSPONDANLKDARLLWPKYKQYKHSIADWLITGNVALTFMGF 170
DB 112 REGASTGTQRFAPLNSPONDANLKDARLLWPKYKQYKHSIADWLITGNVALTFMGF 161
QY 181 KTFSGAGRDWPEEDYNGABETWIGDKRYEGDRELENPGLAVOMGLIYVNPDPNG 240
DB 172 KTFGCGGVNDWHPEDYNGABETWIGDKRYEGDRELENPGLAVOMGLIYVNPDPNG 231
QY 241 KPDPYAAARDIRETFGRMANDEETVALTAGGTFGKTHGAADAEKTVGREPAAGTEEM 300
DB 232 KPDPYAAARDIRETFGRMANDEETVALTAGGTFGKTHGAAGGPA-THVGPPEAPIDEAQ 290
QY 301 SLGNKNTYTGCHGADTTISGLEAWTKPTPONSNNFFENLFGYEWELTKSPAGAYOMKPK 360
DB 291 GLGMSISYSGKGSOTITSGEGAWTPPTOMDTSYFDMLFGDWMLTKSPAGAYOMKPK 350
QY 361 DGAGAGTIPDAHDPSKSHAPFMTDLDALMDPDQYEKISRYYENPDEADAFKAWTKL 420
DB 351 DPDEKDLAPDAEDPSKPYPMNTDLDALFPDPEYETKARHFQNPPEFAFAFAWTKL 410
QY 481 ANASASTFNSDKRGANGARILAPQKDMVNNPQOLARVLTKEIGDIEFNOAQSNK 540
DB 524 KYSLADLIYLAGCAVEKAAKADGHEVQVFPNGRADATAEQTDVDFEFAVLEPPADGRN 583
QY 601 YTKPEKHSYAEMLYDRAQLLSAPETALYGVGMKRVLTNYDGSQIGVFTNKGPSLND 660
DB 584 YKORQSYVPEELLVDKALQGLTGTPMTVLVGLGLVAGNYRDLPHGVFTDRIGLVND 643
QY 661 FVNLIDLNKTNWASDSQKVEGRDFTGCVKNSQTRVDLIFGNSRLALARYGCAD 720
DB 644 FVNLIDLNKTNWYPTDSSG--IYETROKTVGEVMTATRVDLIFGNSILRSTAEFFAQDD 701

```

RESULT 4

T43873

Catalase (BC 1.11.1.6) [imported] - synthetic

C:Species: synthetic

A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*

C:Date: 24-Mar-2000 #sequence\_revision 24-Mar-2000 #text\_change 23-Mar-2001

C:Accession: T43873

K:Katsura, T.; Miyai, K.; Takumaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto

A:Title: Evolutionary engineering by random elongation mutagenesis.

A:Reference number: Z22644; PMID:93115998; PMID:9320270

A:Accession: T43873

A>Status: translated from GB/EMBL/DBJ  
 A:Accession: U751.1CWA>  
 A:Cross-references: EMBL:AB020075; NID:q4140057; PIDN:BAA35987.1; PID:q4140058  
 C:Comment: For the wild-type sequence, see PR:J05020.  
 C:Genetics:  
 A:Gene: cat  
 C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; Db 4; Length 751;  
 Best Local Similarity 68.7%; Pred. No. 5.3e-177;  
 Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

```

QY 1 MENKHSGSYTYWTFGGCKPFTGSGLSQAGGCTKRWMPNKLNLGLRHQSLSDPN 60
DB 1 MENO-----NRONAOCPEHGSVYTNQSS-NRTNKDMWPNQNLSTLHQRDKNPH 51
QY 61 DDPDYAEFFKLDLAAYKDLAALDSDSDWADYGYGTPFFIRMAHSGAGTYRGDC 120
DB 52 DEEFYAEFFKLDLAAYKDLAALDSDSDWADYGYGTPFFIRMAHSGAGTYRGDC 111
QY 121 RGGSGSGSORPAPLNSPONDANLKDARLLWPKYKQYKHSIADWLITGNVALTFMGF 180
DB 112 REGASTGTQRFAPLNSPONDANLKDARLLWPKYKQYKHSIADWLITGNVALTFMGF 171
QY 181 KTFSGAGRDWPEEDYNGABETWIGDKRYEGDRELENPGLAVOMGLIYVNPDPNG 240
DB 172 KTFGCGGVNDWHPEDYNGABETWIGDKRYEGDRELENPGLAVOMGLIYVNPDPNG 231
QY 241 KPDPYAAARDIRETFGRMANDEETVALTAGGTFGKTHGAADAEKTVGREPAAGTEEM 300
DB 232 KPDPYAAARDIRETFGRMANDEETVALTAGGTFGKTHGAAGGPA-THVGPPEAPIDEAQ 290
QY 301 SLGNKNTYTGCHGADTTISGLEAWTKPTPONSNNFFENLFGYEWELTKSPAGAYOMKPK 360
DB 291 GLGMSISYSGKGSOTITSGEGAWTPPTOMDTSYFDMLFGDWMLTKSPAGAYOMKPK 350
QY 361 DGAGAGTIPDAHDPSKSHAPFMTDLDALMDPDQYEKISRYYENPDEADAFKAWTKL 420
DB 351 DPDEKDLAPDAEDPSKPYPMNTDLDALFPDPEYETKARHFQNPPEFAFAFAWTKL 410
QY 421 THRMGPKRYVLGPEVPOEDLIWQDPIPVSHPLVDENDIEGAKLTLESGITVSELVST 480
DB 411 THRMGPKRYVLGPEVPEKDFINQDPIPEVDYELT-EAEIEEIKAKLTLSGLTVSELVST 469
QY 481 ANASASTFNSDKRGANGARILAPQKDMVNNPQOLARVLTKEIGDIEFNOAQSNK 540
DB 470 ANASASTFNSDKRGANGARILAPQKDMVNEPERLAKVSVTDIOREL-----PK 523
QY 541 AVSLADLIYLAGCAVEKAAKADGHEVQVFPNGRADATAEQTDVDFEFAVLEPPADGRN 600
DB 524 KYSLADLIYLAGCAVEKAAKADGHEVQVFPNGRADATAEQTDVDFEFAVLEPPADGRN 583
QY 601 YTKPEKHSYAEMLYDRAQLLSAPETALYGVGMKRVLTNYDGSQIGVFTNKGPSLND 660
DB 584 YKORQSYVPEELLVDKALQGLTGTPMTVLVGLGLVAGNYRDLPHGVFTDRIGLVND 643
QY 661 FVNLIDLNKTNWASDSQKVEGRDFTGCVKNSQTRVDLIFGNSRLALARYGCAD 720
DB 644 FVNLIDLNKTNWYPTDSSG--IYETROKTVGEVMTATRVDLIFGNSILRSTAEFFAQDD 701
QY 721 SEEFKVDYFQKAWAKVMDLDRFL 744
DB 702 NQEFYDFDINAWVYVMDLDRFL 725

```

RESULT 5

T43874

Catalase (BC 1.11.1.6) [imported] - synthetic

C:Species: synthetic

A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*

C:Date: 24-Mar-2000 #sequence\_revision 24-Mar-2000 #text\_change 23-Mar-2001

C:Accession: T43874

R:MaSaura, T.; Miyai, K.; Trakulnaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto, K. 1998. Evolutionary molecular engineering by random elongation mutagenesis. A:Title: Evolutionary molecular engineering by random elongation mutagenesis. A:Reference number: 22644; MUID:95116598; PMID:992070  
A:Accession: T43874  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-751 <MAY>  
A:Cross-references: EMBL:AB020079; NID:94140065; PIDN:BA36991.1; PID:94140066  
C:Keywords: for the wild-type sequence, see PIR:J05020.  
C:Gene: cat  
C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB 4; Length 751;  
Best Local Similarity 68.7%; Pred. No. 5.3e-177;  
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

```

QY 1 MHNKHSSTNTNTGKCPFGGSGAGGCTKNDWPNMLNGLRHQHSLSDN 60
DB 1 MHNQ-----NQNAQCFPGISVYNQSS-RTTNKDWPNMLNLSLHQRDKTNPH 51
QY 61 DDPDYAEFPKLDLAAYKDDLAALMTDSQWMPADYGHGPTFIRMAHSAGTVRGD 120
DB 52 DEENTAEFOKLDLWALKEDLAKLTESQWMPADYGHGPTFIRMAHSAGTVRGD 111
QY 121 RGGSGSORFAPLNSWPDNNDKARLLWPKIKQYGRKISWADLMLTGNVALETMGF 180
DB 112 RGAAGTGORFAPLNSWPDNNDKARLLWPKIKKYGKISWADLFLAGNVALESWG 171
QY 181 KTFGAGRADYWPEDYWGAEYEMJADKRYEGDRELNPGLAVQMLTYNPGPG 240
DB 172 KITFGGGRVYWPEDYWGAEYEMJASERYSDRELENPLAAVQMLTYNPGDG 231
QY 241 KPDPAARADIRETFRGMNDEETVALAGHTGKHAADAERYGPREPAAAGTEEM 300
DB 232 KPDPAARADIRETFRGMNDEETVALAGHTGKHAAGGAGA-THVGPEPEAPTEAQ 290
QY 301 SLGKNTYTGHGADITTSGLGCAWTKPTQWNSNFTFENLFCYEWELTSPAGAYOKPK 360
DB 291 GLWISSYGGKGSDDITTSGLGCAWTKPTQWNTSYFDMFLGYDMLTSPAGAWQWAV 350
QY 421 THRMGPKRYLGPPEVPKEDIMQDTPDYSHPLVDNDIEGLKAKILSGLTYSYELVST 480
DB 411 THRMGPKRYLGPPEVPKEDIMQDTPDYSHPLVDNLT-ENAEIEIKAKILSGLTYSYELVST 469
QY 481 ANASASTFRNSDKGGANGARTILAPKQWENVNPOOLARVLTLEGIQEDFNQAQSDNK 540
DB 470 ANASASTFRNSDKGGANGARTILAPKQWENVNPEERLAKVLSYEDLOREL-----PK 523
QY 541 AVSLADILVLACGAEVKAADGAEVQVYPPNPGADATAGTQDVEAFEPADGPN 600
DB 524 KYSIADILVLGSSAAVEKAARDAGDKVYPPFPGRGDQGTQDVESEFVLEPFDGPN 583
QY 601 TYKPEKYSKEMLVDRAGLLSLSPMTALVSGMRVLTNTDGSORGVFTNKPGQNSD 660
DB 584 YQKEISYVPEELLVQKALGLTAPMTLVGLVGLVGNANTRLDPHGVTGRTIGVLTND 643
QY 661 EYVNLDDNTKWRASDESDKVEGHDKFGVWMSSTRVDLI FGSNSLSALAEVYGCAD 720
DB 644 EYVNLDDNTYEVNPTDSG--IYEIRDRKTEGVMTATRVDLIFGSNSLSASTAEYTAQDD 701
QY 721 SEEFKPVKDFKAWKAVMDLDFDL 744
DB 702 NOEFKPVDFNTAWKYVMAADRFDL 725

```

R:MaSaura, T.; Miyai, K.; Trakulnaleamsai, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto, K. 1998. Evolutionary molecular engineering by random elongation mutagenesis. A:Title: Evolutionary molecular engineering by random elongation mutagenesis. A:Reference number: 22644; MUID:95116598; PMID:992070  
A:Accession: T43874  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-751 <MAY>  
A:Cross-references: EMBL:AB020079; NID:94140065; PIDN:BA36991.1; PID:94140066  
C:Keywords: for the wild-type sequence, see PIR:J05020.  
C:Gene: cat  
C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB 4; Length 751;  
Best Local Similarity 68.7%; Pred. No. 5.3e-177;  
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

```

QY 1 MHNKHSSTNTNTGKCPFGGSGAGGCTKNDWPNMLNGLRHQHSLSDN 60
DB 1 MHNQ-----NQNAQCFPGISVYNQSS-RTTNKDWPNMLNLSLHQRDKTNPH 51
QY 61 DDPDYAEFPKLDLAAYKDDLAALMTDSQWMPADYGHGPTFIRMAHSAGTVRGD 120
DB 52 DEENTAEFOKLDLWALKEDLAKLTESQWMPADYGHGPTFIRMAHSAGTVRGD 111
QY 121 RGGSGSORFAPLNSWPDNNDKARLLWPKIKQYGRKISWADLMLTGNVALETMGF 180
DB 112 RGAAGTGORFAPLNSWPDNNDKARLLWPKIKKYGKISWADLFLAGNVALESWG 171
QY 181 KTFGAGRADYWPEDYWGAEYEMJADKRYEGDRELNPGLAVQMLTYNPGPG 240
DB 172 KITFGGGRVYWPEDYWGAEYEMJASERYSDRELENPLAAVQMLTYNPGDG 231
QY 241 KPDPAARADIRETFRGMNDEETVALAGHTGKHAADAERYGPREPAAAGTEEM 300
DB 232 KPDPAARADIRETFRGMNDEETVALAGHTGKHAAGGAGA-THVGPEPEAPTEAQ 290
QY 301 SLGKNTYTGHGADITTSGLGCAWTKPTQWNSNFTFENLFCYEWELTSPAGAYOKPK 360
DB 291 GLWISSYGGKGSDDITTSGLGCAWTKPTQWNTSYFDMFLGYDMLTSPAGAWQWAV 350
QY 421 THRMGPKRYLGPPEVPKEDIMQDTPDYSHPLVDNDIEGLKAKILSGLTYSYELVST 480
DB 411 THRMGPKRYLGPPEVPKEDIMQDTPDYSHPLVDNLT-ENAEIEIKAKILSGLTYSYELVST 469
QY 481 ANASASTFRNSDKGGANGARTILAPKQWENVNPOOLARVLTLEGIQEDFNQAQSDNK 540
DB 470 ANASASTFRNSDKGGANGARTILAPKQWENVNPEERLAKVLSYEDLOREL-----PK 523
QY 541 AVSLADILVLACGAEVKAADGAEVQVYPPNPGADATAGTQDVEAFEPADGPN 600
DB 524 KYSIADILVLGSSAAVEKAARDAGDKVYPPFPGRGDQGTQDVESEFVLEPFDGPN 583
QY 601 TYKPEKYSKEMLVDRAGLLSLSPMTALVSGMRVLTNTDGSORGVFTNKPGQNSD 660
DB 584 YQKEISYVPEELLVQKALGLTAPMTLVGLVGLVGNANTRLDPHGVTGRTIGVLTND 643
QY 661 EYVNLDDNTKWRASDESDKVEGHDKFGVWMSSTRVDLI FGSNSLSALAEVYGCAD 720
DB 644 EYVNLDDNTYEVNPTDSG--IYEIRDRKTEGVMTATRVDLIFGSNSLSASTAEYTAQDD 701
QY 721 SEEFKPVKDFKAWKAVMDLDFDL 744
DB 702 NOEFKPVDFNTAWKYVMAADRFDL 725

```

Db 702 NOEKFYVDFINAWKYNVWADREFDL 725

RESULT 7  
T43880

Catalase (EC 1.11.1.6) [imported] - synthetic  
C:Species: synthetic  
A:Note: cat gene engineered and expressed in Bacillus stearothermophilus  
C:Accession: T43880  
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.  
A:Reference number: 222644; MUID:93116998; PMID:9320270  
A:Accession: T43880  
A:Species: synthetic  
A:Molecule type: DNA  
A:Residues: 1-751 <MUT>  
A:Cross-references: EMBL:AB020091; NID:94140089; PIDN:BA037003.1; PID:94140090  
C:Comment: For the wild-type sequence, see PIR:J50520.  
C:Genetics:  
A:Gene: cat  
C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB: 4; Length 751;  
Best Local Similarity 68.7%; Pred. No. 5.3e-177;  
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

Qy 1 MENIKHSGSSTYNTGCGCPPTGSGSLKSGAGGCTKRDWPNMLNIGLILQHSSISDPN 60  
Db 1 MENQ-----NRONAOCFFHGSYVWQS-NRTTKDMPNQNLSTLHQRKTPH 51  
Qy 61 DPDFYAEFKLDAIAVKDLAALMTDSQDWMPADYGHYGPFFRMAHSAGTYRIGDG 120  
Db 52 DEEFNFAEFPKLDYALKEOLAKLMTESQDWMPADYGHYGPFFRMAHSAGTYRIGDG 111  
Qy 121 RGGSGSGRFPAPLNSPNDKARLLWPPIKQYGRKISWADLMLTGNVALETMGF 180  
Db 112 RGSAGTGTORFAPLNSPNDKARLLWPPIKQYGRKISWADLMLTGNVALETMGF 171  
Qy 181 KTFGAGRAGVWPEEDYVWGPAGTFLWGLKRYGSDRELENPAGVOMGLIYVNPFGNG 240  
Db 172 KTFGGGRVWHPPEEDYVWGSERLWGLKRYGSDRELENPAGVOMGLIYVNPFGNG 231  
Qy 301 SLGKNTYGTGHCADTITSGLEGAWTKPTQMSNFFENLGYEVELTSPAGAKOMPK 360  
Db 232 KDFKAAANDLRETFRMCHNDDEETVALIAGGTGFKAGGAP-TIVGPEPAPIEAQ 290  
Qy 361 DGACAGTIPDNDHDKPSHAPPLTDLALMDPOYETISRYYENPDPEADAFANAKYL 420  
Db 291 GLGWSITGKSGSDTITSGLEGAWTKPTQMSNFFENLGYEVELTSPAGAKOMPK 350  
Qy 421 THRMGPKRYLGPVPEQDLIMODIPDYSHPLVNDTEGLKAKLTSGSLVSELVST 480  
Db 411 THRMGPKRYLGPVPEQDLIMODIPDYSHPLVNDTEGLKAKLTSGSLVSELVST 469  
Qy 481 ANASASTFNSDKRGANGARILAPQKDWVNNPQGLARVLTLEGIOEDFNOASDNK 540  
Db 470 ANASASTFNSDKRGANGARILAPQKDWVNNPQGLARVLTLEGIOEDFNOASDNK 533  
Qy 541 AVSLADILVLACGACGKAAADGAGHEVQFPNPGRADATAEQTDEAFALPAADGFRN 600  
Db 534 KYSTADILVLGSSAAVEKAAADGAGDVKVPFPFGRGDQTQSDTDESFVLEPDAFGFRN 583  
Qy 601 YIKPKHYSAEMLVDRAQLLSLAPENTKALYGCWKYLGTYNDGSGHQVFTNKPQLSND 660  
Db 584 YQKHSIVYPELLVDRAQLLQATAPMTVVLGGLKRYLGANYRHPGIVFTDRIGLUND 643  
Qy 661 FFWLLDNTWASDESQKPYGGRDPTGCEYKNSGTRVDLIPGNSLSILHQRKTPH 720

Db 644 FFWLLDNTWVPTDGS--IYEHDRKTGEVMTATRVLDIPGNSILRSTAEFTAQDD 701  
Qy 721 SEEFYKDYKYNKAWKYNWADREFDL 744  
Db 702 NOEKFYVDFINAWKYNVWADREFDL 725

RESULT 8  
T43880

Catalase (EC 1.11.1.6) [imported] - synthetic  
C:Species: synthetic  
A:Note: cat gene engineered and expressed in Bacillus stearothermophilus  
C:Accession: T43880  
A:Title: Evolutionary molecular engineering by random elongation mutagenesis.  
A:Reference number: 222644; MUID:93116998; PMID:9320270  
A:Accession: T43880  
A:Species: synthetic  
A:Molecule type: DNA  
A:Residues: 1-751 <MUT>  
A:Cross-references: EMBL:AB020099; NID:94140105; PIDN:BA037011.1; PID:94140106  
C:Comment: For the wild-type sequence, see PIR:J50520.  
C:Genetics:  
A:Gene: cat  
C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB: 4; Length 751;  
Best Local Similarity 68.7%; Pred. No. 5.3e-177;  
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

Qy 1 MENIKHSGSSTYNTGCGCPPTGSGSLKSGAGGCTKRDWPNMLNIGLILQHSSISDPN 60  
Db 1 MENQ-----NRONAOCFFHGSYVWQS-NRTTKDMPNQNLSTLHQRKTPH 51  
Qy 61 DPDFYAEFKLDAIAVKDLAALMTDSQDWMPADYGHYGPFFRMAHSAGTYRIGDG 120  
Db 52 DEEFNFAEFPKLDYALKEOLAKLMTESQDWMPADYGHYGPFFRMAHSAGTYRIGDG 111  
Qy 121 RGGSGSGRFPAPLNSPNDKARLLWPPIKQYGRKISWADLMLTGNVALETMGF 180  
Db 112 RGSAGTGTORFAPLNSPNDKARLLWPPIKQYGRKISWADLMLTGNVALETMGF 171  
Qy 181 KTFGAGRAGVWPEEDYVWGPAGTFLWGLKRYGSDRELENPAGVOMGLIYVNPFGNG 240  
Db 172 KTFGGGRVWHPPEEDYVWGSERLWGLKRYGSDRELENPAGVOMGLIYVNPFGNG 231  
Qy 241 KDPPIAARDLRETFRMCHNDDEETVALIAGGTGFKAGGAP-TIVGPEPAPIEAQ 300  
Db 232 KDFKAAANDLRETFRMCHNDDEETVALIAGGTGFKAGGAP-TIVGPEPAPIEAQ 290  
Qy 301 SLGKNTYGTGHCADTITSGLEGAWTKPTQMSNFFENLGYEVELTSPAGAKOMPK 360  
Db 232 KDFKAAANDLRETFRMCHNDDEETVALIAGGTGFKAGGAP-TIVGPEPAPIEAQ 290  
Qy 361 DGACAGTIPDNDHDKPSHAPPLTDLALMDPOYETISRYYENPDPEADAFANAKYL 420  
Db 351 DPDEKALAPDNDHDKPSHAPPLTDLALMDPOYETISRYYENPDPEADAFANAKYL 410  
Qy 421 THRMGPKRYLGPVPEQDLIMODIPDYSHPLVNDTEGLKAKLTSGSLVSELVST 480  
Db 411 THRMGPKRYLGPVPEQDLIMODIPDYSHPLVNDTEGLKAKLTSGSLVSELVST 469  
Qy 481 ANASASTFNSDKRGANGARILAPQKDWVNNPQGLARVLTLEGIOEDFNOASDNK 540  
Db 470 ANASASTFNSDKRGANGARILAPQKDWVNNPQGLARVLTLEGIOEDFNOASDNK 533  
Qy 541 AVSLADILVLACGACGKAAADGAGHEVQFPNPGRADATAEQTDEAFALPAADGFRN 600  
Db 534 KYSTADILVLGSSAAVEKAAADGAGDVKVPFPFGRGDQTQSDTDESFVLEPDAFGFRN 583

QY 601 YIKPEHKVSAEMVLVDSRAQLLISAPMELVLGSHRVLTNTDGSQHGVTNNKPOQLSND 660  
 Db 584 YKQKESYVPPBELVLVDKALIGLTAPBMVLVGLGLRVIGANTDLPHGVFTDRIQVIND 643  
 QY 661 FYNVLNDLNTKWRASDESQKVEGRDPTKGEVYKMSCTRVDLIFGNSLSRLAALAEVYCAD 720  
 Db 644 FYNVLNDLNTKWRASDESQKVEGRDPTKGEVYKMSCTRVDLIFGNSLSRLAALAEVYCAD 701  
 QY 721 SEEFKVDKFKAWKAVMDLDRFDL 744  
 Db 702 NOEFKVDKFKAWKAVMDLDRFDL 725

RESULT 9  
 T43877  
 catalase (EC 1.11.1.6) [imported] - synthetic  
 C:Species: synthetic  
 A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*  
 C:Date: 24-Mar-2000 #sequence\_revision 24-Mar-2000 #text\_change 23-Mar-2001  
 R:Accession: T43877  
 R:Masuura, T.; Miyai, K.; Takahataleamasa, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto, T.; Biotechnol. 17, 58-61, 1998  
 A:Title: Evolutionary molecular engineering by random elongation mutagenesis.  
 A:Reference: T43877 222644; MUID:95116598; PMID:9520270  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1753 <MAY>  
 A:Cross-references: EMBL:AB020089; NID:94140085; PID:BAA37001.1; PID:94140086  
 C:Comment: For the wild-type sequence, see PIR:J80520.  
 C:Genetics:  
 C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB 4; Length 752;  
 Best Local Similarity 68.7%; Pred. No. 5,3e-177;  
 Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;  
 QY 1 MENKHSSSTVNTNGKCPFGGSLKSGAGGCTKKNRDNWPNMLNGLIRHQHSLSDPN 60  
 Db 1 MENO-----NRQNAQCFPHSGVTNQS-NRTTNKDNWPNQLNLSILHQHDKRNP 51  
 QY 61 DPDFYAEPEKLDLAAYKDLAALMTDQDMPADYGHGPPFIRNAHSHSAGTVRIQDG 120  
 Db 52 DEEFNVAEEFOKLDYALKDELAKLMTESQDMPADYGHGYPFIRNAHSHSAGTVRIQDG 111  
 QY 121 RGGGSGSQRFAPLNSWPNNDNADKARLLWPKIKQYGRKISWADLMLTGNVALTEPMGF 180  
 Db 112 RGAAGTGTQRFAPLNSWPNNDNADKARLLWPKIKKYYGNKISWADLFLAGNVAIESMG 171  
 QY 181 KTFGAGRADYWPEDYVKGAEYEMLDGRYEGDRELNPAGVQGLIYVNPSPG 240  
 Db 172 KTFGGGRVYWPEDYVKGAEYEMLDGRYEGDRELNPAGVQGLIYVNPSPG 231  
 QY 241 KPDTAAARDIRETFGRMANDEETVALTAGCTHGAADAEKYVGRPEAAGTEPM 300  
 Db 232 KPDKAAARDIRETFGRMANDEETVALTAGCTHGAAGAGA-THVGPEPEAPTEAQ 290  
 QY 301 SLGRNVTGTCGADITTSGLGANTKTPTQSHNFFENLFCYEWELETKSPAGYQKPK 360  
 Db 291 GLWISYVKGAGSDITTSGLGANTKTPTQSHNFFENLFCYEWELETKSPAGYQKPK 350  
 QY 361 DGAGACTIPDADHDSKSHAPFLMTDLALRMDPOYEKISRHYENDEPDEADAPAKAYKL 420  
 Db 351 DPDKADLPADHDSKSHAPFLMTDLALRMDPOYEKISRHYENDEPDEADAPAKAYKL 410  
 QY 421 THRMGPKRVYLGPEVPEQDLINQDPIPOVSHPLVDENDIECLKAKILSGLTIVSELVST 480  
 Db 411 THRMGPKRVYLGPEVPEQDLINQDPIPOVSHPLVDENDIECLKAKILSGLTIVSELVST 469  
 QY 481 AMASASTFRNSDKGCCANGARILAPOKDWEVNEPRLARVLTLEGTOEDNQASDNK 540  
 Db 470 AMASASTFRNSDKGCCANGARILAPOKDWEVNEPRLARVLTLEGTOEDNQASDNK 523

QY 541 AVSLADLIVLAGCAVEKAAKADGAEVQVPPNPRADTAETQDVFAEALPEAADGFRN 600  
 Db 524 KVSINDLIVLGSSAVERKAAKADGAEVQVPPNPRADTAETQDVFAEALPEAADGFRN 583  
 QY 601 YIKPEHKVSAEMVLVDSRAQLLISAPMELVLGSHRVLTNTDGSQHGVTNNKPOQLSND 660  
 Db 584 YKQKESYVPPBELVLVDKALIGLTAPBMVLVGLGLRVIGANTDLPHGVFTDRIQVIND 643  
 QY 661 FYNVLNDLNTKWRASDESQKVEGRDPTKGEVYKMSCTRVDLIFGNSLSRLAALAEVYCAD 720  
 Db 644 FYNVLNDLNTKWRASDESQKVEGRDPTKGEVYKMSCTRVDLIFGNSLSRLAALAEVYCAD 701  
 QY 721 SEEFKVDKFKAWKAVMDLDRFDL 744  
 Db 702 NOEFKVDKFKAWKAVMDLDRFDL 725

RESULT 10  
 T43882  
 catalase (EC 1.11.1.6) [imported] - synthetic  
 C:Species: synthetic  
 A:Note: cat gene engineered and expressed in *Bacillus stearothermophilus*  
 C:Date: 24-Mar-2000 #sequence\_revision 24-Mar-2000 #text\_change 23-Mar-2001  
 R:Accession: T43882  
 R:Masuura, T.; Miyai, K.; Takahataleamasa, S.; Yomo, T.; Shima, Y.; Miki, S.; Yamamoto, T.; Biotechnol. 17, 58-61, 1998  
 A:Title: Evolutionary molecular engineering by random elongation mutagenesis.  
 A:Reference: T43882 222644; MUID:95116598; PMID:9520270  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1753 <MAY>  
 A:Cross-references: EMBL:AB020119; NID:94140145; PID:BAA37031.1; PID:94140146  
 C:Comment: For the wild-type sequence, see PIR:J80520.  
 C:Genetics:  
 C:Keywords: oxidoreductase

Query Match 67.8%; Score 2714.5; DB 4; Length 753;  
 Best Local Similarity 68.7%; Pred. No. 5,3e-177;  
 Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;  
 QY 1 MENKHSSSTVNTNGKCPFGGSLKSGAGGCTKKNRDNWPNMLNGLIRHQHSLSDPN 60  
 Db 1 MENO-----NRQNAQCFPHSGVTNQS-NRTTNKDNWPNQLNLSILHQHDKRNP 51  
 QY 61 DPDFYAEPEKLDLAAYKDLAALMTDQDMPADYGHGPPFIRNAHSHSAGTVRIQDG 120  
 Db 52 DEEFNVAEEFOKLDYALKDELAKLMTESQDMPADYGHGYPFIRNAHSHSAGTVRIQDG 111  
 QY 121 RGGGSGSQRFAPLNSWPNNDNADKARLLWPKIKQYGRKISWADLMLTGNVALTEPMGF 180  
 Db 112 RGAAGTGTQRFAPLNSWPNNDNADKARLLWPKIKKYYGNKISWADLFLAGNVAIESMG 171  
 QY 181 KTFGAGRADYWPEDYVKGAEYEMLDGRYEGDRELNPAGVQGLIYVNPSPG 240  
 Db 172 KTFGGGRVYWPEDYVKGAEYEMLDGRYEGDRELNPAGVQGLIYVNPSPG 231  
 QY 241 KPDTAAARDIRETFGRMANDEETVALTAGCTHGAADAEKYVGRPEAAGTEPM 300  
 Db 232 KPDKAAARDIRETFGRMANDEETVALTAGCTHGAAGAGA-THVGPEPEAPTEAQ 290  
 QY 301 SLGRNVTGTCGADITTSGLGANTKTPTQSHNFFENLFCYEWELETKSPAGYQKPK 360  
 Db 291 GLWISYVKGAGSDITTSGLGANTKTPTQSHNFFENLFCYEWELETKSPAGYQKPK 350  
 QY 361 DGAGACTIPDADHDSKSHAPFLMTDLALRMDPOYEKISRHYENDEPDEADAPAKAYKL 420  
 Db 351 DPDKADLPADHDSKSHAPFLMTDLALRMDPOYEKISRHYENDEPDEADAPAKAYKL 410  
 QY 421 THRMGPKRVYLGPEVPEQDLINQDPIPOVSHPLVDENDIECLKAKILSGLTIVSELVST 480  
 Db 411 THRMGPKRVYLGPEVPEQDLINQDPIPOVSHPLVDENDIECLKAKILSGLTIVSELVST 469

Db 411 TERONGKRYTLOPEVKEEDTIPHOQDTPEDVYELT-EAEIEBEIKAKILNSGLTFVSELVY 469  
 Qy 481 ANASASTERNDSKRGANGARIELAPOKDQEVNVPQOLARVLKLTGEOEDENQASDNK 540  
 Db 470 ANASASTERNDSKRGANGARIELAPOKDQEVNPEPERLAKVLSVTEIDQREL-----FK 523  
 Qy 541 AYSIADLIVLAGCAGYKAAKADAGHEVOVFNPCRADATAEOTDVERFEALPEADGFRN 600  
 Db 524 KYSIADLIVLAGCAGYKAAKADAGHEVOVFFPFRDQDQEDQEDVSEFANVLEFADGFRN 583  
 Qy 601 YTKPEHKVSAEMKLDRAQGLLSISAPMETALVGMGRVLGTNDKSGOEVFTNPKQSLND 660  
 Db 594 YKOQEVSVPPBELVLDKQGLLGTAPEMVVLVGLVGLVAGYARLDLPGVFTDRIQVLTND 643  
 Qy 661 FFWMLDLMTKRAESDSKVEGFRGFTGKYGWSTRVOLIFGNSNSELARALAEVTCAD 720  
 Db 644 FFWMLDLMTKRAESDSKVEGFRGFTGKYGWSTRVOLIFGNSNSELARALAEVTCAD 701  
 Qy 721 SEERKFYDQVAKARVNDLDRFDL 744  
 Db 702 NOERKFYDQVAKARVNDLDRFDL 725

RESULT 11  
 catalase/peroxidase XF2232 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 24-Aug-2001  
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing, 131-137, 2000  
 A:Note: This is the complete list of the plant pathogen Xylella fastidiosa.  
 A:Reference number: AB2515; MUID:20355717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: F82584  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-781 <SIM>  
 A:Experimental source: strain 9a5c  
 R:Simpton, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S., submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm, C.; Junqueira, M.A.; Kemp, E.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; Laig, C.; Martins, E.M.P.; Matsukuma, A.; Marinho, C.; Marques, M.; Martins, E.M.P.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sanlelli, R.V.; Sawalak, A.; da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zaccaria, M., 2002  
 C:Contents: annotation  
 A:Gene: XF2232  
 C:Superfamily: catalase HPI  
 F:126/Active site: His (distal axial ligand) #status predicted  
 F:313/Binding site: heme iron (His) (proximal axial ligand) #status predicted  
 F:356/Active site: Trp (tryptophyl radical intermediate) #status predicted

Query Watch 67.6% Score 2706.5; DB 2; Length 781;  
 Best Local Similarity 67.1%; Pred. No. 2e-176;  
 Matches 510; Conservative 86; Mismatches 131; Indels 33; Gaps 7;

Qy 9 SSTVNTNGKCPFTGCSIKOSAGGKTKNDWPNMLGLIRQHSLSDPNDPFDYAE 68  
 Db 30 ASSETTANVKKP-----NTKATVSTHKNWPNQKLVLLRQHSNKNPLGFTFYAK 84  
 Qy 69 EFKKDLAAVKDLAALWTSDDWADPAGYGVGFTFIMAHWSAGTYRGDGGCGSGS 128  
 Db 85 EFKKDLAAVKDLAALWTSDDWADPAGYGVGFTFIMAHWSAGTYRGDGGCGAG 144

Qy 129 QRFAPLNSPNNAKDKARLLIIPKQYKRIKSWADLMTLGNVALPMGFKTSPAG 188  
 Db 145 QRFAPLNSPNDVSLDKARLLIIPKQYKRIKSWADLMTLGNVALPMGFKTSPAG 204  
 Qy 189 RADVNPREDVYAGKTEML--GDKRY-- --BQ-----DRELENPLG 224  
 Db 205 RADVNPREDVYAGKTEML--GDKRY-- --BQ-----DRELENPLG 224  
 Qy 225 AYVNGLLIYVNPGRGPDPTAAAGDRIETFORAMNDRETVALLIGITFKTIGAGDA 284  
 Db 265 AYVNGLLIYVNPGRGPDPTAAAGDRIETFORAMNDRETVALLIGITFKTIGAGDA 324  
 Qy 285 EKYVGREFAAGATENSLGKNTYGTGHGADTITSLSGAWTKTPTQNSNFFENLGE 344  
 Db 325 DM-VGPEFAGELDQGLGKINRFGSGKAGDTITSLSGAWTKTPTQNSNFFENLGE 383  
 Qy 345 WELTKSPAGYQWYKWDGAGATGIDADPSKSHAPFMTLDTALRMDPDEKISRYE 404  
 Db 384 WELTKSPAGYQWYKWDGAGATGIDADPSKSHAPFMTLDTALRMDPDEKISRYE 441  
 Qy 405 NFDEFAFAKARYKLTTRMDGPKRYILGPEVPOEDLIWDQDTPIDVSHPLVDENDI 464  
 Db 442 HPDQFAFAKARYKLTTRMDGPKRYILGPEVPOEDLIWDQDTPIDVSHPLVDENDI 501  
 Qy 465 AKTILESGLTSELSTANASASTERNDSKRGANGARIELAPOKDQEVNVPQOLARV 524  
 Db 502 QKTSASGLTISQVLSSTANASASTERNDSKRGANGARIELAPOKDQEVNVPQOLARV 561  
 Qy 525 LKGIQEDFNQAGSNKAVSLADLIVLAGCAGYKAAKADAGHEVOVFPENGRADAT 584  
 Db 562 LKRVGTTFN-AQAGDKRISEADLIVLAGGVGVQAAKRAQTVVPEVPGRTDALOR 620  
 Qy 585 VEAEPEALPEADGCFRNTYEPBKVSAEMKLDRAQGLLSISAPMETALVGMGRVL 644  
 Db 621 VSSPAPLEPPADGCFRNTYEPBKVSAEMKLDRAQGLLSISAPMETALVGMGRVL 680  
 Qy 645 SOHCVFTNKPQSLNSDFVNLNLDKRAESDSKVEGFRGFTGKYGWSTRVOLIFG 704  
 Db 701 VHGVTQRTQSLNSDFVNLNLDKRAESDSKVEGFRGFTGKYGWSTRVOLIFG 740  
 Qy 765 SNSELARALAEVTCADSEKPKVDFYKAAKQVNDLDRFDL 744  
 Db 741 SNSELARALAEVTCADSEKPKVDFYKAAKQVNDLDRFDL 780

RESULT 12  
 catalase/peroxidase [imported] - Caulobacter crescentus  
 C:Species: Caulobacter crescentus  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eلسen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, P.J.; Kholmavac, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C., 2001  
 A:Title: Complete Genome Sequence of Caulobacter crescentus.  
 A:Reference number: AB7249; MUID:21173698; PMID:11259647  
 A:Accession: AB7249  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-737 <STO>  
 A:Cross references: GB:AE005673; NID:g13424687; PIDN:ANK25005.1; GSPDB:GN00148  
 C:Gene: CC3043  
 C:Superfamily: catalase HPI

Query Watch 67.2% Score 2691; DB 2; Length 737;  
 Best Local Similarity 69.5%; Pred. No. 2.1e-175;  
 Matches 508; Conservative 64; Mismatches 145; Indels 14; Gaps 6;

Qy 15 NYGCKCPFTGCSIKOSAGGKTKNDWPNMLGLIRQHSLSDPNDPFDYAEFKLD 74  
 Db 15 NYGCKCPFTGCSIKOSAGGKTKNDWPNMLGLIRQHSLSDPNDPFDYAEFKLD 74





```

QY 251 TBTFGRMANDEETVALLIAGHTTGTGHCADAEKXYVGREPAAGIENSLGWANTYGT 310
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 245 IRATFGNNGNDEETVALLIAGHTTGLKXGAA-AAASHVGADPEAAPTEAQCLGWASSYGS 303
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 311 GHGADYITISLEGAWTKTPTONSNFFENLFGYENELTKSPAGAYQWKPDKGAGAGTIPD 370
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 GVGADAITSGLEVWVOTPTONSNFFENLFGYENELTKSPAGAYQWKPDKGAGAGTIPD 361
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 371 ANDPSKSHAPMLTDTALRMDPDYKISRYENDEFADAFKAWKLTHERMDGPKYR 430
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 362 PEDFSKKRKTMLVTDLTLREDPEPEKISRRELPDQAFNEAFARAFWFKLTHERMDGPKAR 421
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 431 YLGSFVPOEDLIWQDTPDVNSHPLVD--ENDIEGLKAKILESGLTVSELSTAWASASTF 488
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 422 YIGFVPEKDLIWQDPLP---QELYQPTQEDINLKAAIAASGLSISEMVSVAWASASTF 478
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 489 RNSDKGCGANGARTLAPKQKMYANPOOLARVLKTEGLOEDNOMQSDNKAYSIADLI 548
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 479 RGGDKGCGANGARLALAPQDMEYN--AVALRVLPLVLEALQXTNKA-----SLADI 529
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 549 VLAGCAGYKAAKAGHEVQVFPNFGADATABQTDVEAFEALEPAADCFRNYIKPEHKV 608
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 530 VLAGVWGLEQAAAAGYSVSPFAGVQARQDQTDIEMFSLPEPTADGFRNYRARLWYS 589
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 609 SAEPLVDRAQLLSNAPETALYGGHVLCTYKGSORGVFTNKCOLSNDFVNLIDL 668
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 590 TTESLLIDKAAQLTLTAPMTVLVGGHVLCTYKGSORGVFTNKCOLSNDFVNLIDL 649
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 669 NTKWRASDESKVFGRDKTGEVKNWSTGYDLIFGNSSELRALAEVYGCADSEKFKVD 728
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 650 RYEMKPTDANELFEGORLGTGEVYATRADLVFGNSVLRALAEVYACSDAHEKFKVD 709
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 729 FYANAKYMDLORFDL 744
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 710 FYAANYKVMNLOREDL 725
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: October 7, 2003, 19:46:46  
 Job time : 32.3425 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 7, 2003, 19:38:38 : Search time 74.5995 Seconds  
(all elements)  
2577.085 Million cell updates/sec

Title: US-09-884-889-8

Perfect score: 4002  
Sequence: 1 MENEKSSSTNTNGKRC.....VKDFVAKWVGLDREDLK 745

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23.\*  
1: sp-archaea.\*  
2: sp-archaeal.\*  
3: sp-fungi.\*  
4: sp-human.\*  
5: sp-invertebrate.\*  
6: sp-mammal.\*  
7: sp-nbce.\*  
8: sp-organelle.\*  
9: sp-phage.\*  
10: sp-plant.\*  
11: sp-rodent.\*  
12: sp-virus.\*  
13: sp-vertebrate.\*  
14: sp-unclassified.\*  
15: sp-virus.\*  
16: sp-bacterioph.\*  
17: sp-archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2989.5	74.7	736	17 Q8TSC4	Q8TSC4 methanosa
2	2884.5	72.1	756	16 Q987S0	Q987S0 ribitolium
3	2867	71.6	748	2 Q939D2	Q939D2 burkholderi
4	2787.5	69.7	727	16 Q92XG8	Q92XG8 ribitolium
5	2762.5	69.0	756	16 Q8PM44	Q8PM44 xanthomonas
6	2753.5	68.8	735	16 Q9KEE6	Q9KEE6 bacillus ha
7	2715.5	67.8	743	2 Q9SE80	Q9SE80 bacillus st
8	2714.5	67.8	743	2 Q9SSR0	Q9SSR0 bacillus st
9	2714.5	67.8	744	2 Q9SSN2	Q9SSN2 bacillus st
10	2714.5	67.8	744	2 Q9SSK7	Q9SSK7 bacillus st
11	2714.5	67.8	745	2 Q9SSQ5	Q9SSQ5 bacillus st
12	2714.5	67.8	746	2 Q9SSN4	Q9SSN4 bacillus st
13	2714.5	67.8	748	2 Q9SSL8	Q9SSL8 bacillus st
14	2714.5	67.8	749	2 Q9SSP6	Q9SSP6 bacillus st
15	2714.5	67.8	749	2 Q9SSQ4	Q9SSQ4 bacillus st
16	2714.5	67.8	750	2 Q9SSQ7	Q9SSQ7 bacillus st

17	2714.5	67.8	751	2 Q9SSP7	Q9SSP7 bacillus st
18	2714.5	67.8	751	2 Q9SS15	Q9SS15 bacillus st
19	2714.5	67.8	751	2 Q9SSM6	Q9SSM6 bacillus st
20	2714.5	67.8	751	2 Q9SSP1	Q9SSP1 bacillus st
21	2714.5	67.8	751	2 Q9SSL1	Q9SSL1 bacillus st
22	2714.5	67.8	751	2 Q9SSM4	Q9SSM4 bacillus st
23	2714.5	67.8	751	2 Q9SSM1	Q9SSM1 bacillus st
24	2714.5	67.8	751	2 Q9SSQ8	Q9SSQ8 bacillus st
25	2714.5	67.8	751	2 Q9SSP9	Q9SSP9 bacillus st
26	2714.5	67.8	751	2 Q9SSM3	Q9SSM3 bacillus st
27	2714.5	67.8	751	2 Q9SSN3	Q9SSN3 bacillus st
28	2714.5	67.8	751	2 Q9SSM0	Q9SSM0 bacillus st
29	2714.5	67.8	751	2 Q9SSK4	Q9SSK4 bacillus st
30	2714.5	67.8	751	2 Q9SSP2	Q9SSP2 bacillus st
31	2714.5	67.8	751	2 Q9SSQ0	Q9SSQ0 bacillus st
32	2714.5	67.8	751	2 Q9SSM6	Q9SSM6 bacillus st
33	2714.5	67.8	751	2 Q9SSM7	Q9SSM7 bacillus st
34	2714.5	67.8	751	2 Q9SSM7	Q9SSM7 bacillus st
35	2714.5	67.8	751	2 Q9SSM7	Q9SSM7 bacillus st
36	2714.5	67.8	751	2 Q9SSM8	Q9SSM8 bacillus st
37	2714.5	67.8	751	2 Q9SSL6	Q9SSL6 bacillus st
38	2714.5	67.8	751	2 Q9SSL9	Q9SSL9 bacillus st
39	2714.5	67.8	751	2 Q9SSL3	Q9SSL3 bacillus st
40	2714.5	67.8	751	2 Q9SSQ1	Q9SSQ1 bacillus st
41	2714.5	67.8	751	2 Q9SSQ2	Q9SSQ2 bacillus st
42	2714.5	67.8	751	2 Q9SSQ2	Q9SSQ2 bacillus st
43	2714.5	67.8	751	2 Q9SSP8	Q9SSP8 bacillus st
44	2714.5	67.8	751	2 Q9SSM2	Q9SSM2 bacillus st
45	2714.5	67.8	751	2 Q9SSN8	Q9SSN8 bacillus st

## ALIGNMENTS

### RESULT 1

ID	Q8TSC4	PRELIMINARY:	PRT: 736 AA.
AC	Q8TSC4;		
DT	01-JUN-2002 (TrEMBLrel. 21, Created)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 21, Last annotation update)		
DE	Catalase/peroxidase.		
GN	MA0972.		
OS	Methanosarcina acetivorans.		
OC	Archaea; Euryarchaeota; Methanococci; Methanosarcinales;		
OC	Methanosarcinaceae; Methanosarcina.		
OX	NCBI_TaxID=2214;		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CJA / ATCC 35395 / DSM 2834;		
RX	MEDLINE=21929760; PubMed=11932238;		
RA	Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,		
RA	Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,		
RA	Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,		
RA	Linton L., McSwan P., McKernan K., Talamas J., Tirrell A., Ye W.,		
RA	Hedderich P., Bartram-Smitthi I., Graham D.E., Graham D.A., Guss A.M.,		
RA	Leigh J.A., Li W., Liu J., Mukhopadhyay B.C., Nussbaum R., Smith K.,		
RA	Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,		
RA	Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,		
RA	Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,		
RA	Metcalfe W., Birren B.;		
RT	"The genome of Methanosarcina acetivorans reveals extensive metabolic		
RT	and physiological diversity."		
RT	EMBL: A010742; GenBank: AF040021;		
RT	EMBL: A010742; GenBank: AF040021;		
DR	InterPro: IPR000763; Bac_catalase/pease.		
DR	InterPro: IPR002016; peroxidase.		
DR	Pfam: PF00141; peroxidase; 1.		
DR	TIGRfams: TIGR00198; cat_per_HPI; 1.		
DR	PROSITE: PS00435; PEROXIDASE; 1.		
DR	PROSITE: PS00436; PEROXIDASE; 2; 1.		
DR	Peroxidase; Complete proteome.		

SQ	SEQUENCE	736 AA; 81051 MW; 67A0508CD91BFA7A CRC64;			
	Query Match	74.7%; Score 2889.5; DB 17; Length 736;			
	Best Local Similarity	76.3%; Pred. No. 9e-185;			
	Matches	551; Conservative	62; Mismatches	108; Indels	1; Gaps
QY	23	TGGSLSKSGAGCTKNDKMNPNMLGILGKHSSLSLSDPNPDYDAEERFKLDAALVKDL	82		
DB	15	TSGANKETGSDMSNDWNLKLELHQBHSSKSNPMGEDFNTAKEKSLDAALVKDL	74		
QY	83	PALMTDSQWADYCHVGFPTIRMAHWSAGTYVGGSGGSGSOREPAPYNSNDAN	142		
DB	75	AALMTDSQWADYCHVGFPTIRMAHWSAGTYVGGSGGSGSOREPAPYNSNDAN	134		
QY	143	LDAKALLMPTIKQYGRKISWADMLTGNVALETMGFTKTFAGRADWPEDEYVWG	202		
DB	135	LDAKALLMPTIKQYGRKISWADMLTGNVALETMGFTKTFAGRADWPEDEYVWG	194		
QY	203	AETNHEVGEYGDRELEPLGVNOMGLIYVNEPCGPKGPDPTAAARDIETPCRWAND	262		
DB	195	SEDTWLDGERTGDRLEPLGVNOMGLIYVNEPCGPKGPDPTAAARDIETPCRWAND	254		
QY	263	EETVALTAGHTFGKTHGAADAERYGPREPAAAGTEMSLGRNRYTGSGADTISGLE	322		
DB	255	EETVALTAGHTFGKTHGAADAERYGPREPAAAGTEMSLGRNRYTGSGADTISGLE	313		
QY	323	GAMTKTTPQWSNFFENFICFWEMLTSPAGYQYKPDGAGAGTIPDAHDPKSHAPN	382		
DB	314	VTTTPTKMSNFFRILFCFWEMLTSPAGYQYKPDGAGAGTIPDAHDPKSHAPN	373		
QY	383	LTDLALMDPDYKISRRYENPDEPDAFAKAWYKLTHRDMPKRYLGPPEPORDI	442		
DB	374	LTDLALMDPDYKISRRYENPDEPDAFAKAWYKLTHRDMPKRYLGPPEPORDI	433		
QY	443	WQDPTPOVSLPVDNDLGLAKILLESGLTVELSVLSTAWASASTFRNSKGGANGARI	502		
DB	434	WQDPTPAVNHLEIDKZAFPLKDRILLASGLISQVSTAWASASTFRNSKGGANGARI	493		
QY	503	PLAPQKQVNNPQOLARVLKTLGQTDENPQASDKWASVADLTLVLAGCAVERAND	562		
DB	494	PLAPQKQVNNPQOLARVLKTLGQTDENPQASDKWASVADLTLVLAGCAVERAND	553		
QY	563	AGHEVQVFPGRADAFATQDVEAPALBPADGFRTNPKYHKSVAESMLVDAQLLS	622		
DB	554	AGDYVTFPLPGRMDALQDQSGVFPALPQADGFRTNPKYHKSVAESMLVDAQLLS	613		
QY	622	LSAPRTALVGGNRYLGTNDSQHGFTYNNKPOLSNDFPYNLLDNTKWRASDESQYF	682		
DB	614	LSAPRTALVGGNRYLGTNDSQHGFTYNNKPOLSNDFPYNLLDNTKWRASDESQYF	673		
QY	683	EGRDFKTEYKMSGTVDLIFGNSRLAALAEYGCATSEKFKYKDFKAWAKVNDLDR	742		
DB	674	EGCDKRYGKMYGTVDLIFGNSRLAALAEYGCATSEKFKYKDFKAWAKVNDLDR	733		
QY	743	DL 744			
DB	734	DL 735			

RP	SEQUENCE FROM N.A.				
RC	STRAIN-MAFF303099				
XX	NCBI-TaxID=2102950; pubMed=11214969;				
BA	Kato T, Asamizu E, Y. Saito S, Asamizu E, Kato T, Sasamoto S.,				
RA	Watanabe A, Ideawata K., Ishikawa K., Kawashima K., Kimura T.,				
RA	Kishida Y., Kiyokawa K., Kohara M., Matsumoto M., Matsuno A.,				
RA	Mochizuki Y., Nakayama S., Kohazaki N., Shimpō S., Sugimoto M.,				
RA	Takeuchi C., Tamada M., Tabata S.;				
RT	*Complete genome structure of the nitrogen-fixing symbiotic bacterium				
RT	Mesorhizobium loti, strain 38400.				
DB	NCBI-TaxID=38400;				
DB	EMBL: AF0030310; AF553110.1; -				
HSP:	P00431; 1BS; Bac.ctase/prase.				
DR	InterPro: IPR00763; Bac.ctase/prase.				
DR	InterPro: IPR002016; Peroxidase.				
DR	PFam: PF00141; peroxidase.1.				
DR	PRINTS: PR00458; PEROXIDASE.1.				
DR	TIGRFAM5: TIGR00198; cat_per_HPI.1.				
DR	PROSITE: PS00436; PEROXIDASE.2.1.				
DR	PROSITE: PS00436; PEROXIDASE.2.1.				
DR	Peroxidase; Complete proteome.				
KW	SEQUENCE 756 AA; 82348 MW; 788BAFC5DC5900AC CRC64;				
SQ	SEQUENCE 756 AA; 82348 MW; 788BAFC5DC5900AC CRC64;				
	Query Match	72.1%; Score 2884.5; DB 16; Length 756;			
	Best Local Similarity	71.7%; Pred. No. 5.8e-178;			
	Matches	533; Conservative	74; Mismatches	125; Indels	7; Gaps
QY	2	ENIKHSKSGSTYNTWTKGCFPGGSLKAGGGCTKNDWPNMLNLGLTQHSLSLSDPD	61		
DB	19	ONSSSDMGARTDNDNSAGKCPVARGS-----AGRTNRDMPNQLNYQLAQSSLSDPMG	72		
QY	62	PDYDAEERFKLDAALVKDLAALMTDSQWADYCHVGFPTIRMAHWSAGTYVWG	121		
DB	73	EADYAEERFKLDAALVKDLAALMTDSQWADYCHVGFPTIRMAHWSAGTYVWG	132		
QY	122	GGGSGSGSOREPAPYNSNDANLDAKALLMPTIKQYGRKISWADMLTGNVALETMGFK	181		
DB	133	GGAGAGQORFAPYNSNDANLDAKALLMPTIKQYGRKISWADMLTGNVALETMGFK	192		
QY	182	TFGAGRADWPEDEYVWGAGTETKDKRYEGDRELEPLGVNOMGLIYVNEPCGPKG	241		
DB	193	TFGAGRADWPEDEYVWGAGTETKDKRYEGDRELEPLGVNOMGLIYVNEPCGPKG	252		
QY	242	PDYDAEERFKLDAALVKDLAALMTDSQWADYCHVGFPTIRMAHWSAGTYVWG	301		
DB	253	PDYDAEERFKLDAALVKDLAALMTDSQWADYCHVGFPTIRMAHWSAGTYVWG	311		
QY	302	LQWAKYTGKAGADTITSGLEGNATKTPQWSNFFENFICFWEMLTSPAGYQYKPDG	361		
DB	312	LQWAKYTGKAGADTITSGLEGNATKTPQWSNFFENFICFWEMLTSPAGYQYKPDG	371		
QY	362	GAGAGTIPDAHDPKSHAPNLTDLALMDPDYKISRRYENPDEPDAFAKAWYKLT	421		
DB	372	GAGAGTIPDAHDPKSHAPNLTDLALMDPDYKISRRYENPDEPDAFAKAWYKLT	431		
QY	422	HRMGPVARYLGPVPEKELPQDVPYVYDVLIDEDQDAEALKAELTASGLSVSLVSTA	481		
DB	432	HRMGPVARYLGPVPEKELPQDVPYVYDVLIDEDQDAEALKAELTASGLSVSLVSTA	491		
QY	482	WASATFRNSKGGANGARIAPQKDEYNPQOLARVLKTLGQTDENPQASDKRKA	541		
DB	492	WASATFRNSKGGANGARIAPQKDEYNPQOLARVLKTLGQTDENPQASDKRKA	551		
QY	542	YSLADLTLVGGNAGTEKAAKAAGHSVDFPFWPMDASOBOTDTHSFAPLEPVDGFRNY	601		
DB	552	YSLADLTLVGGNAGTEKAAKAAGHSVDFPFWPMDASOBOTDTHSFAPLEPVDGFRNY	611		
QY	602	IKPEKYSVAEMLVDRAPQOLLSLAPENTALVGGNRYLGTNDSQHGFTYNNKPOLSNDF	661		
DB	612	YSSKQRTVYEMKLVDRAPQOLLSLAPENTALVGGNRYLGTNDSQHGFTYNNKPOLSNDF	671		
QY	662	PNYLLDNTKWRASDESQYFEGDRELEPLGVNOMGLIYVNEPCGPKGFTVQHSLSLSDPD	721		
DB	721	PNYLLDNTKWRASDESQYFEGDRELEPLGVNOMGLIYVNEPCGPKGFTVQHSLSLSDPD			

RESULT 2

Q98750	PRELIMINARY; PRT; 756 AA.
Q98750	
AD	Q98750
DT	01-OCT-2001 (TrEMBLrel. 18, Created)
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE	Catalase/peroxidase.
GN	M16940.
OS	Rhizobium loti (Mesorhizobium loti).
CS	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC	Proteobacteria; Rhizobiales; Mesorhizobium.
OX	NCBI_TaxID=38400.
RN	[1]

Db 672 FVNLDTGTEKATSDAKVDFEGRDKRYGVKWTGTRADLLFGSISGSLALAEVATADA 731  
 Qy 732 BEKEVDFVYKAWKVDLDRDL 744  
 Db 732 KAKFADKVFVYKAWKVDLDRDL 754

RESULT 3  
 ID Q939D2 PRELIMINARY; PRT; 748 AA.  
 AC Q939D2: 2001 (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TEMBLrel. 23, Last annotation update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Catalase-peroxidase protein katG.  
 OS Burkholderia pseudomallei (Pseudomonas pseudomallei).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Burkholderia.  
 OX NCBI\_TaxID=28450;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RA "Soprasant S., Sallabhan R., Whangsek W., Mongkolsuk S.;  
 RA "Burkholderia pseudomallei, oxygen mutant has altered biofilm  
 RA "formation and extracellular protease activity, and  
 RT oxidative stress response".  
 RL Submitted (SSEP-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AY040244; AAK72466.3; -  
 DR HSP; P00431; I8ES.  
 DR InterPro; IPR002016; Peroxidase.  
 DR Pfam; PF00441; peroxidase\_1.  
 DR PROSITE; PS00435; PEROXIDASE\_1.  
 DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
 DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
 KW Peroxidase.  
 SQ SEQUENCE 748 AA; 81760 MW; F59D6C85A6850AF3 CRC64;

Query Match  
 Similarity 71.0%; Pred. No. le-171;  
 Matches 541; Conservative 64; Mismatches 112; Indels 14; Gaps 5;

Qy 19 KCPTFGSLQSGAGGTTKRWPNKNGILRHSSISLSDPNDPDYAEFFKLLAAV 78  
 Db 26 KCFF-----HQANGSTNSRNDWPNQDLSILRHSSISLSDPKNDNTAQAFKLLAAV 80  
 Qy 79 KQDLAALMTSDQKADYGVGHYPTFMAMHSAGCTYRIGDGRGGSGSRFAPLNSWP 138  
 Db 81 KQDLAALMTSDQKADYGVGHYPTFMAMHSAGCTYRIGDGRGGSGSRFAPLNSWP 140  
 Qy 139 DNANLDRKLLWPIKQYGRKLSWADLLTGNVALETGKFTFGAGRDWVEED 198  
 Db 141 DNANLDRKLLWPIKQYGRKLSWADLLTGNVALETGKFTFGAGRDWVEED 199  
 Qy 199 VTKATFNL-----GDKYEDRELNPLGAVOMGLIYVNPDPGKPDPIAARDIRE 253  
 Db 200 VTKATFNL-----GDKYEDRELNPLGAVOMGLIYVNPDPGKPDPIAARDIRE 259  
 Qy 254 TFGSMANDDETVALTACGFTGCTGAGDAEKYVGPAAACETEMSLGKMTYGTGHC 313  
 Db 260 TFGSMANDDETVALTACGFTGCTGAGDASN-VGASPEAAGTEAQGLSKYRTFGK 318  
 Qy 314 ADTITLSEGNATKPTQNSNFENLFGYVELTETSPAGYQWPFDSAGACTIPDAD 373  
 Db 319 ADATITLSEGVNTTPTQNSNFENLFGYVELTETSPAGYQWPFDSAGACTIPDAD 376  
 Qy 374 PSKSHAPFMITDIALNDPDYKISRYRYENPDPEFADAFANWKLTHRDMDGPKRYLG 433  
 Db 377 PSKKHPFMITDLSLRFDPAYEKISRYRYENPDPEFADAFANWKLTHRDMDGPKRYLG 436  
 Qy 434 RPYQPDVILNDPDYEDVSLVNDMDTEGLKAKLLESGITVVELSVTAMASSTRNSDK 493  
 Db 437 PEVPAEVLNDPDYEDVSLVNDMDTEGLKAKLLESGITVVELSVTAMASSTRNSDK 496

Qy 494 RGAGCARIRLAPQKDMNEVNPQOLARVATKLEGIQEDFNOAGSDNKAVSLADLVLAGC 553  
 Db 497 RGAGCARIRLAPQKDMNEVNPQOLARVATKLEGIQEDFNOAGSDNKAVSLADLVLAGC 556  
 Qy 554 AGVEKAADKAGCHEVQVNPGRADATQDYEAFALPEADGPRNYLKPHEKYSAREM 613  
 Db 557 AGVQAAKNAHNAVTPFAPGRADASQDQDVEASHVLEPVDGFRNYLKGKRYPAEVL 616  
 Qy 614 LVDRAGLISLAPENTALVGGHVRVLTGNYDQSGHGVFTNKPQOLSDFFVMLDLNKNR 673  
 Db 617 LVDRAGLISLAPENTVVLGGVRLVLAQVQSGHGVFTNKPQOLSDFFVMLDLNKNR 676  
 Qy 674 ASDSDKVEGRDFTGCVKMSGRVLDLIPGNSSELRLAALVYVGCADSEKFKYDVKAM 733  
 Db 677 PTAADVDFGRDRTGELKLTGTRVDLVPGSHSQLRALAAYVYVSADAQEKFRVDFVAVM 736  
 Qy 734 ATWMDVDFDL 744  
 Db 737 NKNVNLDRFDL 747

RESULT 4  
 ID Q92XG8 PRELIMINARY; PRT; 727 AA.  
 AC Q92XG8: 2001 (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE Catalase/peroxidase (EC 1.11.1.6).  
 GN RAL286 OR SMA2379.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid pSyma (megaplasmid 1).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RP STRAIN=1021.  
 RC MEDLINE=2139509; PubMed=11481432;  
 RA Barette, J., Fisher, R.F., Jones, T., Kemp, C., Abola, A.P.,  
 RA Gurjal, M., Hong, A., Huisar, T., Hyman, R.W., Kahle, D., Kaur, M.,  
 RA Kalman, S., Keating, D.H., Palm, C., Peck, M.C., Surzycki, R., Wallis, D.H.,  
 RA Yeh, K.-C., Davis, R.W., Federspiel, N.A., Long, S.R.,  
 FT "Nucleotide sequence and predicted functions of the entire  
 FT "plasmid pSyma (megaplasmid 1) of the bacterium Rhizobium meliloti.  
 FT "Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).  
 DR HSP; P00431; I8ES.  
 DR InterPro; IPR00763; Bac. ctase/prase.  
 DR InterPro; IPR002016; Peroxidase.  
 DR Pfam; PF00141; peroxidase\_1.  
 DR PRINTS; PR00458; PEROXIDASE.  
 DR TIGRFAMs; TIGR00198; cat\_pex\_HPI; 1.  
 DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
 DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
 KW Peroxidase; Sinorhizobium meliloti; Complete proteome.  
 SQ SEQUENCE 727 AA; 80292 MW; 31086D8C7AA793DF CRC64.

Query Match  
 Similarity 69.7%; Score 2787.5; DB 16; Length 727;  
 Best Local Similarity 71.1%; Pred. No. le-171;  
 Matches 521; Conservative 72; Mismatches 127; Indels 13; Gaps 6;

Qy 14 TWTGCGTPTGSLQKSGAGGTTKRWPNKNGILRHSSISLSDPNDPDYAEFFKLL 73  
 Db 5 SSSAGKCP-----VAHTAPGSRNDWPNQDLSILRHSSISLSDPNTYAEFFKLL 59  
 Qy 74 DLAAVKKDLAALMTSDQKADYGVGHYPTFMAMHSAGCTYRIGDGRGGSGSRFAP 133  
 Db 60 DLDAALDRKLLWPIKQYGRKLSWADLLTGNVALETGKFTFGAGRDWVEED 119  
 Qy 134 LNSPQNALDRKLLWPIKQYGRKLSWADLLTGNVALETGKFTFGAGRDWVEED 193





Db 232 KDPKAAARADIRETFRRGNDEETVALIAGSTHFGKAGGAPGATHVGPPEPAAPTEAQ 290  
Qy 301 SLGKNTYTGIGADITISLEGATWTTQTSNFFENIFGYEMELTKSPAGATOWKPK 360  
Db 291 GLOMISSTYGGKGDITISLEGATWTTQTSNFFENIFGYEMELTKSPAGATOWKPK 360  
Qy 361 DCACAGTTPANDDSKSHAPNKTJDLANDMDPEYKISRRYENPDDEPAFAKAWKYL 420  
Db 351 DPDEKPLAPDADPSKVKYPTMMWTTDLALDFEYKISRRYENPDDEPAFAKAWKYL 420  
Qy 421 THRMGPKVRYLGPEVQEDLIMQDPIPOVSHPLVDENDIEGLKAKILESGLTVSELVST 480  
Db 411 THRMGPKVRYLGPEVQEDLIMQDPIPOVSHPLVDENDIEGLKAKILESGLTVSELVST 480  
Qy 481 ANASASTFRNSDKGGANGARTLAPQKDFVNNPOOLARVLTLEGTOEDFNQASDNK 540  
Db 470 ANASASTFRNSDKGGANGARTLAPQKDFVNNPOOLARVLTLEGTOEDFNQASDNK 540  
Qy 541 AVSLADLVLVLAGCAVEKAAKADGHEVQVFPNPGADATAGTQVDEAFEALEPAADGFRN 600  
Db 524 KYSTADLVLVLAGCAVEKAAKADGHEVQVFPNPGADATAGTQVDEAFEALEPAADGFRN 600  
Qy 601 VIKPEKYSAPFEMVDPBAILLSISAPMTALVGVNRYGVTNDOGSQVYFNKPGQSLND 660  
Db 584 YOKQEYSVPPPELLVDKALIGLTAPMTVLVGLRVLGARVLDLPHGVFTDRIGVLTND 643  
Qy 661 FVNLIDNTKWRASDESDFEGKDFKGVKWSGTRVDLIFGNSLSALAEVYTCAD 720  
Db 644 FVNLIDNTKWRASDESDFEGKDFKGVKWSGTRVDLIFGNSLSALAEVYTCAD 720  
Qy 721 SEKFKVDFVYKAWKAWKOLDFDL 744  
Db 702 NOEKVDFDFTNAWKVMNAADFDL 725

RESULT 8  
Q53RQ PRELIMINARY; PRT: 743 AA.  
AC Q53RQ  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).  
GN CAT:illus stearothermophilus  
OC Bacteria: Firmicutes; Bacillales; Geobacillus.  
ON NCBI\_TaxID=1422;  
RX [1]  
RA Matsumura T., Miyai K., Trakunaleamsai S., Yomo T., Shima Y.,  
RA Miki S., Yamamoto K., Urabe I.;  
RA Evolutionary molecular engineering by random elongation  
RT Nat. Biotechnol. 17:58-61(1998).  
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL  
CC -1- CATALYTIC ACTIVITY: 2 H(2O(2)) - O(2) + 2 H(2O).  
CC PEROXIDASE/CATALASE SUBFAMILY.  
DR EMBL: AB020120; BAA37032.1; -.  
DR HSPR: P00431; 1809.  
DR Interpro: IPR00069; Bac\_ctase/prase.  
DR Pfam: PF00141; peroxidase.  
DR PRINTS: PR00458; PEROXIDASE.  
DR TIGRFAMs: TIGR00198; cat\_per\_HPI.1.  
DR PROSITE: PS00435; PEROXIDASE\_1; 1.  
DR PROSITE: PS00436; PEROXIDASE\_2; 1.  
DR Heme; Hydrogen peroxidase; Iron; Oxidoreductase; Peroxidase.  
SQ SEQUENCE 743 AA; 84078 MW; A2C34225A1C5B6F CRC64;  
Query Match 67.8%; Score 2714.5; DB 2; Length 743;  
Best Local Similarity 68.7%; Pred. No. 5,5e-167;  
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

Qy 1 MNKHSSSTNTWTGCKPFGGSLKQAGGTFKNDKPNMJKGTLQHSLSLDPN 60  
Db 1 MNQ-----NRNNAOCCFFGSGVYTNSS-NRTTNDQWPNQLSLIHDNRKTFPH 51  
Qy 61 DPDFVYAEFKLLDLAAVAKLMTDSQWNPADYGYGPPFIRMAHWSACTYRIGDG 120  
Db 120 DEEFYAEFOFKLLDLAKLMTDSQWNPADYGYGPPFIRMAHWSACTYRIGDG 111  
Qy 121 RGGGSGSFRAPLANSFPDNNNDKARLLAMPKQYGRKISWADLMILGNVALPMSG 180  
Db 112 RGGAGSTGTQREAPLANSFPDNNNDKARLLAMPKQYGRKISWADLMILGNVALPMSG 171  
Qy 181 KTFQAGGADRVMPEDYVYGAZEMDLKRYEGTDELENPLGAVOMGLTYVNPQPG 240  
Db 172 KTFQGGGRVYMPEDYVYGAZEMDLKRYEGTDELENPLGAVOMGLTYVNPQPG 231  
Qy 241 KPDPTAARADIRETFRRGNDEETVALIAGSTHFGKAGGAPGATHVGPPEPAAPTEAQ 290  
Db 232 KDPKAAARADIRETFRRGNDEETVALIAGSTHFGKAGGAPGATHVGPPEPAAPTEAQ 290  
Qy 301 SLGKNTYTGIGADITISLEGATWTTQTSNFFENIFGYEMELTKSPAGATOWKPK 360  
Db 291 GLOMISSTYGGKGDITISLEGATWTTQTSNFFENIFGYEMELTKSPAGATOWKPK 360  
Qy 361 DCACAGTTPANDDSKSHAPNKTJDLANDMDPEYKISRRYENPDDEPAFAKAWKYL 420  
Db 351 DPDEKPLAPDADPSKVKYPTMMWTTDLALDFEYKISRRYENPDDEPAFAKAWKYL 420  
Qy 421 THRMGPKVRYLGPEVQEDLIMQDPIPOVSHPLVDENDIEGLKAKILESGLTVSELVST 480  
Db 411 THRMGPKVRYLGPEVQEDLIMQDPIPOVSHPLVDENDIEGLKAKILESGLTVSELVST 480  
Qy 481 ANASASTFRNSDKGGANGARTLAPQKDFVNNPOOLARVLTLEGTOEDFNQASDNK 540  
Db 470 ANASASTFRNSDKGGANGARTLAPQKDFVNNPOOLARVLTLEGTOEDFNQASDNK 540  
Qy 541 AVSLADLVLVLAGCAVEKAAKADGHEVQVFPNPGADATAGTQVDEAFEALEPAADGFRN 600  
Db 524 KYSTADLVLVLAGCAVEKAAKADGHEVQVFPNPGADATAGTQVDEAFEALEPAADGFRN 583  
Qy 601 VIKPEKYSAPFEMVDPBAILLSISAPMTALVGVNRYGVTNDOGSQVYFNKPGQSLND 660  
Db 584 YOKQEYSVPPPELLVDKALIGLTAPMTVLVGLRVLGARVLDLPHGVFTDRIGVLTND 643  
Qy 661 FVNLIDNTKWRASDESDFEGKDFKGVKWSGTRVDLIFGNSLSALAEVYTCAD 720  
Db 644 FVNLIDNTKWRASDESDFEGKDFKGVKWSGTRVDLIFGNSLSALAEVYTCAD 701  
Qy 721 SEKFKVDFVYKAWKAWKOLDFDL 744  
Db 702 NOEKVDFDFTNAWKVMNAADFDL 725

RESULT 9  
Q53RQ PRELIMINARY; PRT: 744 AA.  
AC Q53RQ  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).  
GN CAT:illus stearothermophilus  
OC Bacteria: Firmicutes; Bacillales; Geobacillus.  
ON NCBI\_TaxID=1422;  
RX [1]  
RA Matsumura T., Miyai K., Trakunaleamsai S., Yomo T., Shima Y.,  
RA Miki S., Yamamoto K., Urabe I.;  
RA Evolutionary molecular engineering by random elongation  
RT Nat. Biotechnol. 17:58-61(1998).  
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL  
CC -1- CATALYTIC ACTIVITY: 2 H(2O(2)) - O(2) + 2 H(2O).  
CC PEROXIDASE/CATALASE SUBFAMILY.  
DR EMBL: AB020120; BAA37032.1; -.  
DR HSPR: P00431; 1809.  
DR Interpro: IPR00069; Bac\_ctase/prase.  
DR Pfam: PF00141; peroxidase.  
DR PRINTS: PR00458; PEROXIDASE.  
DR TIGRFAMs: TIGR00198; cat\_per\_HPI.1.  
DR PROSITE: PS00435; PEROXIDASE\_1; 1.  
DR PROSITE: PS00436; PEROXIDASE\_2; 1.  
DR Heme; Hydrogen peroxidase; Iron; Oxidoreductase; Peroxidase.  
SQ SEQUENCE 743 AA; 84078 MW; A2C34225A1C5B6F CRC64;  
Query Match 67.8%; Score 2714.5; DB 2; Length 743;  
Best Local Similarity 68.7%; Pred. No. 5,5e-167;  
Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

CC -1- CATALYTIC ACTIVITY: 2 H(2O)(2) - O(2) + 2 H(2)O  
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL.  
 CC PEROXIDASE/CATALASE SUBFAMILY.  
 DR EMBL: AB020092; BAA37004.1; -;  
 DR HSP: P00431; 1b39.  
 DR InterPro: IPRO00763; Bac\_ctase/prase.  
 DR InterPro: IPRO02016; Peroxidase.  
 DR PRISM: PR00415; Peroxidase 1.  
 DR TRIGRAMS: TIGR00198; cat\_per\_HPI; 1.  
 DR PROSITE: P500435; PEROXIDASE\_1; 1.  
 DR PROSITE: P500436; PEROXIDASE\_2; 1.  
 DR Heme: Hydrogen peroxide; Iron; Oxidoreductase; peroxidase.  
 KW SEQ: SEQUENCE 744 AA; 83946 MW; 34632D2B5B1095B CRC64;

Query Match 67.8%; Score 2714.5; DB 2; Length 744;  
 Best Local Similarity 68.7%; Pred. No. 5-157;  
 Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

Dh 1 MENKHSGSYTYNTGCKPFTGGSLQKAGGCTKNRWPMNUNLILGHSHSLSDPN 60  
 1 MENO-----RQNAACQCPHGSVTNQSS-NRTTNKQMPNQLNLSLHODRKTNP 51  
 61 DPDPVAFBEFKLDAVKDLAALNTSDQWADYGHGPPFTIRMAHISAGTYRIGD 120  
 52 DEEFNAYEFPKLDYALKEDRLKLTESODWADYGHGPPFTIRMAHISAGTYRIGD 111  
 121 RGGGSSQSFAPLNSPNDNKLARLLWPKYGRKISWADMLITGNVALETMGF 180  
 112 RGGAGTGTORFAPLNSPNDNKLARLLWPKYGRKISWADMLITGNVALETMGF 171  
 181 KTFPGAGGADWVEEDYTCAGTWMGLDKRYEGDRENLPLAVQNGLYTVNPGSDG 240  
 172 KTIGFGGGRVDWHEEDYTVGSEKWLASERYSGDRENLPLAVQNGLYTVNPGSDG 231  
 241 KPDPFAAADIRRETFGRMANDEETVALIAGGTGKTHGAADAASKYVREPAAGIEM 300  
 232 KPDPFAAADIRRETFGRMANDEETVALIAGGTGKTHGAADAASKYVREPAAGIEM 290  
 301 SLGNKNTYGTGADTITSGLEGAWTKTPQSNFNFENLGYEVELTSPAGAYQWPK 350  
 291 GLGNTISSYKGGKSGSTITSGLEGAWTPPTQMDTSTFDMLFGDMLNLTSPAGAYQW 360  
 361 DGAGAGTIPDAHPDQSKSHAFPLMTDIALRHDPDYETKISRYYENPDEFAFAKAYKL 420  
 351 DPDEKOLAPDAEDPSKYPTMMTDLARDFDETEKIAHFQAFPEAFARAWFKL 410  
 421 THRDGPKRYVLGPVEQEDLWQDIPDVSHPDLVDENDIGLKALETSGLYTVSELVST 480  
 411 THRDGPKRYVLGPVEKDFWQDIPDVEYELT--EATIEETKAKLTNSGLTVSELVST 469  
 481 AWASATFNRSKRGKGANGARILAPQDKMEVNNPQOLARLKTLEGIOEDFNAQSDNK 540  
 470 AWASATFNRSKRGKGANGARILAPQDKMEVNEPEKARLYKLVSYEDIQEL-----PK 523  
 541 AVSLADITVLACAGCAKKAADGHEVQDPNPGADAKATARDYVFAFLPADQPRN 600  
 534 KYVSLADITVLACAGCAKKAADGHEVQDPNPGADAKATARDYVFAFLPADQPRN 583  
 601 YIKPEHYSAEMLVDRAGLSLAPMETALYQGMVRLGTVNYSQIGVFNKPOLNSD 660  
 584 YQKISVYPPBEILLVDRAGLQALAPMETALYQGMVRLGTVNYSQIGVFNKPOLNSD 643  
 661 FVYVLLDNLNTWASDSKVEFGDFKTEVKSSTGRVLDLFGSNEALRALAYTCAD 720  
 644 FVYVLLDNLNTWASDSKVEFGDFKTEVKSSTGRVLDLFGSNEALRALAYTCAD 701  
 721 SEEFYKDFVFAWAKVMDLDFD 744  
 702 NQKFAFYDFVFAWAKVMDLDFD 725

RESULT 10

Q9SSK7 PRELIMINARY; PRT; 744 AA.  
 AC Q9SSK7;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).  
 GN CAT.  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillales; Geobacillus.  
 CX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Matsutera T., Miyai K., Trakunaleeamsai S., Yomo T., Shima Y.,  
 RA Miki S., Yamamoto K., Urabe I.;  
 RA "Evolutionary molecular engineering by random elongation  
 mutagenesis", vol. 17, 59-61 (1998).  
 BL -1- CATALYTIC ACTIVITY: 2 H(2O)(2) - O(2) + 2 H(2)O  
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL  
 CC PEROXIDASE/CATALASE SUBFAMILY.  
 CC EMBL: AB020067; BAA36979.1; -;  
 DR HSP: P00431; 1b39.  
 DR InterPro: IPRO00763; Bac\_ctase/prase.  
 DR InterPro: IPRO02016; Peroxidase.  
 DR PRISM: PR00414; Peroxidase 1.  
 DR TRIGRAMS: TIGR00198; cat\_per\_HPI; 1.  
 DR PROSITE: P500435; PEROXIDASE\_1; 1.  
 DR PROSITE: P500436; PEROXIDASE\_2; 1.  
 KW Heme: Hydrogen peroxide; Iron; Oxidoreductase; Peroxidase.  
 KW SEQ: SEQUENCE 744 AA; 84000 MW; 35CED214D4F6AB CRC64;

Query Match 67.8%; Score 2714.5; DB 2; Length 744;  
 Best Local Similarity 68.7%; Pred. No. 5-157;  
 Matches 511; Conservative 75; Mismatches 139; Indels 19; Gaps 6;

Dh 1 MENKHSGSYTYNTGCKPFTGGSLQKAGGCTKNRWPMNUNLILGHSHSLSDPN 60  
 1 MENO-----RQNAACQCPHGSVTNQSS-NRTTNKQMPNQLNLSLHODRKTNP 51  
 61 DPDPVAFBEFKLDAVKDLAALNTSDQWADYGHGPPFTIRMAHISAGTYRIGD 120  
 52 DEEFNAYEFPKLDYALKEDRLKLTESODWADYGHGPPFTIRMAHISAGTYRIGD 111  
 121 RGGGSSQSFAPLNSPNDNKLARLLWPKYGRKISWADMLITGNVALETMGF 180  
 112 RGGAGTGTORFAPLNSPNDNKLARLLWPKYGRKISWADMLITGNVALETMGF 171  
 181 KTFPGAGGADWVEEDYTCAGTWMGLDKRYEGDRENLPLAVQNGLYTVNPGSDG 240  
 172 KTIGFGGGRVDWHEEDYTVGSEKWLASERYSGDRENLPLAVQNGLYTVNPGSDG 231  
 241 KPDPFAAADIRRETFGRMANDEETVALIAGGTGKTHGAADAASKYVREPAAGIEM 300  
 232 KPDPFAAADIRRETFGRMANDEETVALIAGGTGKTHGAADAASKYVREPAAGIEM 290  
 301 SLGNKNTYGTGADTITSGLEGAWTKTPQSNFNFENLGYEVELTSPAGAYQWPK 350  
 291 GLGNTISSYKGGKSGSTITSGLEGAWTPPTQMDTSTFDMLFGDMLNLTSPAGAYQW 360  
 361 DGAGAGTIPDAHPDQSKSHAFPLMTDIALRHDPDYETKISRYYENPDEFAFAKAYKL 420  
 351 DPDEKOLAPDAEDPSKYPTMMTDLARDFDETEKIAHFQAFPEAFARAWFKL 410  
 421 THRDGPKRYVLGPVEQEDLWQDIPDVSHPDLVDENDIGLKALETSGLYTVSELVST 480  
 411 THRDGPKRYVLGPVEKDFWQDIPDVEYELT--EATIEETKAKLTNSGLTVSELVST 469  
 481 AWASATFNRSKRGKGANGARILAPQDKMEVNNPQOLARLKTLEGIOEDFNAQSDNK 540  
 470 AWASATFNRSKRGKGANGARILAPQDKMEVNEPEKARLYKLVSYEDIQEL-----PK 523











**THIS PAGE BLANK (USPTO)**